Comparative genome analysis of *Yersinia*

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften der Fakultät für Biologie der
Ludwig-Maximilians-Universität München

vorgelegt von

Andrey Golubov

aus

Schakhty

Februar 2005

Dissertation eingereicht am: 24.02.2005

Erstgutachter: Prof. Dr. Elisabeth Weiss
Zweitgutachter: Prof. Dr. Martin Parniske
Sondergutachter: Prof. Dr. Dr. Jürgen Heesemann

Tag der mündlichen Prüfung: 06.07.2005

TABLE OF CONTENTS

A. INTRODUCTION

1. *Yersinia* species – general properties

The yersiniae (genus XI of the family *Enterobacteriaceae*) consist of eleven species of which *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica* are considered to be primary pathogens of mammals (Brubaker, 1991). *Y. pseudotuberculosis* and *Y. pestis* are closely related species that share nearly 97% gene homology (Achtman *et al*., 1999; Motin *et al*., 2002; Trebesius *et al*., 1998). *Y. enterocolitica* in contrast presents a more variable genomic arrangement with only 60 - 65% DNA homology with *Y. pestis*/*Y. pseudotuberculosis* (Bottone, 1999).

Today, isolated cases of *Y. pestis* infection (the plague) are reported sporadically in the US, India and Madagascar (Perry and Fetherston, 1997). *Y. pestis* is an obligate parasite, in contrast to *Y. enterocolitica* and *Y. pseudotuberculosis*, which are free-living microorganisms and are food-borne pathogens (Cornelis *et al*., 1998; Black *et al*., 1978). *Y. enterocolitica*, which is the most prevalent in humans, and *Y. pseudotuberculosis* (mainly isolated from animals such as pigs) cause a broad range of gastrointestinal syndromes.

Of special importance to the pathogenic process of all *Yersinia* species is the shared requirement of a virulence plasmid pCD1 (pYV in enteropathogenic *Yersinia*) that encodes a type III secretion system (Cornelis and Van Gijsegem, 2000), which is responsible for injecting into host cells a number of cytotoxins and effectors (*Yersinia* outer proteins) that inhibit bacterial phagocytosis and processes of innate immunity (Brubaker, 2003; Cornelis, 2002). Two additional plasmids unique to *Y. pestis*, termed pPla (pPCP1) (9.6 kb) and pFra (pMT1) (102 kb), play roles in tissue invasion (Lahteenmaki *et al*., 1998) and capsule formation (Kutyrev *et al*., 1986), as well as infection of the plague flea vector (Hinnebusch *et al*., 2002; Hinnebusch, 2003), respectively (Table 1).

The medically significant yersiniae can multiply on appropriate media at temperatures ranging from about 5 to 42°C. However, marked differences mediated by global regulatory mechanisms occur upon an increase from room (26°C) to host (37°C) temperature. These dysfunctions include expression of additional nutritional requirements and production of virulence functions. In contrast to *Y. pestis* which is non-motile, the two enteropathogenic *Yersinia* species are motile at 27 °C (Cover and Aber, 1989; Bottone, 1999; Tauxe *et al*., 1987).

Table 1. Distinguishing properties and virulence determinants of wild-type *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica* (Brubaker, 1991 with modifications).

Gene product and location of genes	Established or putative virulence function[c]	*Y. pestis*	*Y. pseudotuberculosis*	*Y. enterocolitica*
pPla plasmid		+	-	-
Pesticin	-	+	-	-
Plasminogen activator	+	+	-	-
Posttranslational degradation of Yops[a]	+	+	-	-
pCD/pYV plasmid		+	+	+
Yops	+	+	+	+
YadA (protein 1 or Yop A)	+	-	+	+
V antigen	+	+	+	+
pFra plasmid		+	-	-
Fraction 1 or capsular antigen	+	+	-	-
Phospholipase D	+	+	-	-
Chromosomal determinants				
Pigmentation or peptide F (hemin storage at 26°C)	+	+	-	-
Motility (26°C)	-	-	+	+
Hydrophobic sugars in LPS (26°C)/O-antigen	-	-	+	+
Assimilation of low levels of NH_4 (26°C)	-	-	+	+
Constitutive glyoxylate bypass	-	+	-	-
Aspartase	-	-	+	+
Glucose 6-phosphate dehydrogenase	-	-	+	+
Urease	-	-	+	+
Ornithine decarboxylase	-	-	-	+
Host cell invasins				
Invasin (Inv)	+	-	+	+

Attachment invasion locus (Ail)	+	-	-	+
pH 6 antigen/Myf	+	+	+	+
Antigen 5 (catalase)	+	+	+	-
Fermentation of				
Rhamnose	-	-	+	-
Melibiose	-	-	+	-
Sucrose	-	-	-	+
Sorbitol	-	-	-	+
Cellobiose	-	-	-	-
Biosynthesis of				
Methionine	-	-	+	+
Phenylalanine	-	-	+	+
Threonine-glycine	-	-	+	+
Isoleucine-valine	-	-	+	+

[a] +, present; -, absent.

1.1 *Yersinia enterocolitica*

Y. enterocolitica, which is one of the focuses of this study, is widely distributed in nature in aquatic and animal reservoirs, with swine serving as a major reservoir for human pathogenic strains. The species *Y. enterocolitica* was established in 1980 by applying DNA relatedness studies and phenotypic characteristics. A species was defined as that its strains must have a DNA-DNA relatedness of more than 70%. This standard is still valid for the genus Yersinia today. In the past years it was noted that the species *Y. enterocolitica* consists biochemically and serologically heterogeneous strains: so called “European” and “American” biogroups (BG) and serotypes (ST) named after the continent of their first isolation. Isolates of both 16S rRNA gene types had sequence identities of more than 97%. However, it was demonstrated the presence of three DNA-DNA relatedness groups within the species *Y. enterocolitica* represented by the “American” bio- and serotypes, by the enteropathogenic “European” strains and by the non-enteropathogenic “European” strains. Considering the presence of three relatedness clusters and the “minor but consistent phenotypic variation” i.e. the highly conserved 16S rRNA gene sequence of European and American isolates, the division of the species *Y. enterocolitica* into two subspecies was justified. It was proposed the names *Y. enterocolitica* subsp. *enterocolitica* for strains belonging to the 16S rRNA gene type of American origin and *Y. enterocolitica* subsp. *palearctica* for strains belonging to the 16S rRNA gene type of European origin (Neubauer et al., 1999).

Within *Y. enterocolitica*, there exists sufficient biochemical heterogeneity to have warranted the establishment of six biogroups – known as 1A, 1B, 2, 3, 4 and 5 – that can be differentiated by biochemical tests (Table 2) (Bottone, 1999).

Table 2. Biochemical tests used to biogroup *Y. enterocolitica* strains.

Test	Biogroup reaction[a]					
	1A	1B[b]	2	3	4	5
Lipase activity	+	+	-	-	-	-
Salicin (acid production in 24 h)	+	-	-	-	-	-
Esculin hydrolysis (24 h)	+/-	-	-	-	-	-
Xylose (acid production)	+	+	+	+	-	v
Trehalose (acid production)	+	+	+	+	+	-
Indole production	+	+	v	-	-	-
Ornithine decarboxylase	+	+	+	+	+	+(+)
Voges-Proskauer test	+	+	+	+	+	+(+)
Pyrazinamidase activity	+	-	-	-	-	-
Sorbitol (acid production)	+	+	+	+	+	-
Inositol (acid production)	+	+	+	+	+	+
Nitrate reduction	+	+	+	+	+	-

[a] +, positive; -, negative; (+), delayed positive; v, variable.

[b] Biogroup 1B is comprised mainly of strains isolated in the United States.

The latter is further subdivided into three groups: a non-pathogenic group (BG 1A); a weakly pathogenic group that is unable to kill mice (BG 2 to 5); and a high pathogenic, mouse-lethal group (BG 1B) (Carniel, 2002). BG 1A lacks the *Yersinia* virulence plasmid pYV and a functional *inv* gene and seems to be distantly related to the other biogroups. Serologically, *Y. enterocolitica* may be separated into approximately 60 serotypes (ST) of which only 11 serotypes have been most frequently associated with human infection (Table 3). Of these serotypes, the preponderance of infections on a worldwide basis are caused by serotypes O:3, O:9, O:5,27 with a declining number of ST O:8 isolations being made from symptomatic patients. As, however, *Y. enterocolitica* O:3, O:9, and O:8 antigens have been recovered from different *Yersinia* species (Aleksic, 1995) the pathogenic potential of a *Y. enterocolitica* isolate should be based on both serotype and biotype determination (Bottone, 1999). In terms of geographical distribution, the weakly and high pathogenic *Y. enterocolitica* species exhibit some preferences: the high-pathogenic organisms are more frequently isolated in the US,

while the weakly pathogenic isolates are predominantly isolated in Europe and Japan (Aleksic and Bockemuhl, 1990). However, recently *Y. enterocolitica* ST O:8 has been isolated in Germany (Schubert *et al*., 2003). Interestingly, the majority of nonporcine *Y. enterocolitica* isolates are of the non-pathogenic biogroup 1A and lack the virulence determinants of invasive isolates. Furthermore, as shown in Table 3 (Bottone, 1999), there is a close association of pathogenic species with a particular animal reservoir.

Table 3. Virulence of *Y. enterocolitica* correlated with biogroup, serogroup, and ecologic and geographic distribution (Bottone, 1999).

Associated with human infections	Biogroup	Serotype(s)	Ecologic distribution
Yes	1B	O:8, O:4, O:13a,13b, O:18, O:20, O:21	Environment, pigs (O:8), mainly in the United States, Japan, Europe
Yes	2	O:9, O:5,27	Pigs, Europe (O:9), United States (O:5,27), Japan (O:5,27), Sweden, The Netherlands
Yes	3	O:1,2,3, O:5,27, O:3	Chinchilla (O:1,2,3), pigs (O:5,27)
Yes	4	O:3	Pigs, Europe, United States, Japan, South Africa, Scandinavia, Canada, The Netherlands
Yes	5	O:2,3	Hare, Europe
No[a]	1A	O:5, O:6,30, O:7,8, O:18, O:46, nontypeable	Environment, pigs, food, water, animal and human feces, global

[a] *Y. enterocolitica* isolates comprising biogroup 1A may be opportunistic pathogens in patients with underlying disorders.

1.2 *Yersinia pestis*

Yersinia pestis, the causative agent of plague, is supposed to be a recently emerged pathotype evolved from enteropathogen *Y. pseudotuberculosis* (Achtman *et al*., 1999; Skurnik *et al*., 2000). Three *Y. pestis* biovars, Antiqua (A), Mediaevalis (M), and Orientalis (O), that are believed to be the causative agents of the historical plague pandemics, are distinguished by their ability to ferment glycerol and their nitrification activity (Devignat, 1951). However, in addition to these strains there exists a group of *Y. pestis* isolates distributed through various countries of the former USSR, Mongolia, China, and Morocco that share certain characteristics with the closely related species *Y. pseudotuberculosis* (Anisimov *et al*., 2004).

They ferment rhamnose (Rha), are additionally dependent on the nutrients, and possess elective virulence being less virulent in guinea pigs but highly virulent in mice. These strains are described in the literature as cause of occasional human or animal plague cases but have been rarely associated with epizootics of plague (Anisimov *et al*., 2004; Bakanidze *et al*, 2003). To separate these Rha-positive isolates from the main group of *Y. pestis* strains, they were proposed to be named *Y. pestoides* or Pestoides (Martinevskii, 1969). Alternatively they were named according to the place of their first isolation i.e. *Y. pestis* subsp. *caucasica, ulegeica, altaica*, etc. (Anisimov *et al*., 2004).

The main acquisitions of the plague microbe supposed to be responsible for its virulence are two plasmids, in addition to the pYV plasmid. The 10-kb plasmid pPla (also designated as pYP, pPCP1 or pPst) encodes the plasminogen activator and the bacteriocin pesticin. The 100-kb plasmid pFra (also designated pMT1 or pYT) is responsible for the synthesis of Fraction 1 antigen and phospholipase D. The plasminogen activator is involved in the dissemination of the plague bacterium from the site of the initial fleabite, while phospholipase D (previously denoted as murine toxin) plays a major role in survival of plague bacteria in fleas (Hinnebusch *et al*., 2002) . All pathogenic yersiniae share the virulence-associated pYV plasmid, which encodes a finely tuned type III secretion machinery of anti-phagocytic factors (Cornelis and Wolf-Watz, 1997).

Most of the rhamnose-positive *Y. pestis* isolates contain all three *Y. pestis* specific plasmids. However, some of them lack the pPla plasmid and/or carry an enlarged pFra (Filippov *et al*., 1990). *Y. pestis* subsp. *caucasica* (also Pestoides F) is frequently isolated in high mountainous Caucasus and mountainous Dagestan. It belongs phenotypically to biovar Antiqua with *Microtus arvalis* being its main reservoir (Bakanidze *et al*., 2003). Plague epizootics of varying intensity were documented in this focus. Rha-positive *Y. pestis caucasica* strains lack the pPla but contain an enlarged pFra. They are of low virulence in guinea pigs. However, an aerozolised Pestoides F strain lacking the plasminogen activator was proven to be highly virulent (Worsham and Roy, 2003). Strict geographical isolation in a high mountainous region might have led to the preservation of an “ancient plague” microbe. *Y. pestis* G8786 which was isolated in the Caucasian high mountainous focus was typed as an atypical Antiqua strain by genome-wide microarray analysis (Hinchliffe *et al*., 2003). This analysis reflects its remote origin and the highest divergence from other *Y. pestis* strains.

1.3 Pathogenic factors of *Yersinia*

Historically, *Y. enterocolitica* is primarily a gastrointestinal tract pathogen with, under defined host conditions, a strong propensity for extraintestinal spread. When it infects the gastrointestinal tract, acute enteritis with fever and inflammatory, occasionally bloody, watery diarrhea is the most frequent occurrence, particularly in children (Bottone, 1997). In young adults, acute terminal ileitis and mesenteric lymphadenitis mimicking appendicitis appear to be a more common clinical syndrome (Black *et al*., 1978; Chandler and Parisi, 1994). The yersinial nuance in the extent of gastrointestinal tract pathologic findings centers largely about the serogroup of the invading strain, with serogroup O:8 producing the more catastrophic events, including extensive ulceration of the gastrointestinal tract and death (Gutman *et al*., 1973), whereas serogroups O:3 (Lee *et al*., 1990) and O:9 are less destructive in the gastrointestinal tract. Also secondary immunologically mediated sequelae of acute *Y. enterocolitica* infection such as arthritis and erythema nodosum, which are the most common, and Reiter's syndrome, glomerulonephritis, and myocarditis have been reported predominantly among Scandinavians and in the setting of *Y. enterocolitica* serotype O:3, biotype 4 infections. Most patients manifesting postyersinial reactive arthritis are HLA-B27 positive. The reasons underscoring this predisposition are unknown (Bottone, 1997).

Yersinia pestis is primarily a rodent pathogen, with humans being an accidental host when bitten by an infected rat flea. The flea draws viable *Y. pestis* organisms into its intestinal tract. These organisms multiply in the flea and block the flea's proventriculus.

Some *Y. pestis* in the flea are then regurgitated when the flea gets its next blood meal thus transferring the infection to a new host. While growing in the flea, *Y. pestis* loses its capsular layer. Most of the organisms are phagocytosed and killed by the polymorphonuclear leukocytes in the human host. A few bacilli are taken up by tissue macrophages. The macrophages are unable to kill *Y. pestis* and provide a protected environment for the organisms to synthesize their virulence factors. The organisms then kill the macrophage and are released into the extracellular environment, where they resist phagocytosis (YopH and YopE; *Yersinia* outer membrane protein) by the polymorphs. The *Y. pestis* quickly spread to the draining lymph nodes, which become hot, swollen, tender, and hemorrhagic. This gives rise to the characteristic black buboes responsible for the name of this disease.

Within hours of the initial flea bite, the infection spills out into the bloodstream, leading to involvement of the liver, spleen, and lungs. The patient develops a severe bacterial pneumonia, exhaling large numbers of viable organisms into the air during coughing fits. 50 to 60 percent of untreated patients will die if untreated. As the epidemic of bubonic plague

develops (especially under conditions of severe overcrowding, malnutrition, and heavy flea infestation), it eventually shifts into a predominately pneumonic form, which is far more difficult to control and which has 100 percent mortality.

Pathogenicity of *Yersinia* species is mediated by an above mentioned 75 kb plasmid (pYV) and 2 additional plasmids in *Y. pestis* (pPla and pFra). These plasmids control the 4 major virulence factors in the genus: excreted antiphagocytic proteins (Yops), proteins involved in processing and excretion of the Yops (Ysc) and a complex regulatory network (Lcr proteins for low calcium response). Chromosomal factors also play a role in virulence. These factors are the adhesion/invasion proteins (YadA), invasin InvA, the adhesive factor Myf (mucoid *Yersinia* factor, homolog of the pH6 antigen), the enterotoxin Yst, and proteins involved in iron acquisition. Virulence genes on *Yersinia* are controlled by 2 independent regulatory systems, temperature and cellular concentration of calcium. Secretion of Yops occurs only at 37ºC and in the absence of calcium (Cornelis, 1994). A chromosomal gene *ymoA*, encoding YmoA, functions as the temperature controller (Cornelis *et al*., 1991), by blocking VirF protein to the promoters of *yop* genes at temperatures below 30°C and allows expression of *yop* and *ysc* genes at 37°C (Michiels *et al*., 1991). Recently work has been shown that RovA, a member of the MarR/SlyA family of winged-helix transcription factors, regulates expression of *inv* (Revell and Miller, 2000; Ellison *et al*., 2004). Disruption of *rovA* increases the LD_{50} of the organism when inoculated using the oral route. However, when administered by intraperitoneal injection only a slight difference in LD_{50} between mutant and wild-type organisms is apparent.

1.3.1 Yst Enterotoxin

Invasion of *Y. enterocolitica* and *Y. pseudotuberculosis* through mucosal surfaces and colonization of these surfaces is based, at least in part, on the production of a toxin (Yst), a 30-amino acid peptide, similar to the heat stable toxin produced by *E. coli* (Iriarte and Cornelis, 1995). Both toxins activate the particulate form of guanylate cyclase, thus increasing cGMP levels in the intestine. The activity of the toxin eventually leads to fluid accumulation in the intestine. This activity led to the hypothesis that Yst is involved in inducing diarrhea during *Yersinia* infection. Diarrhea is not a major symptom in the mouse infection model; however, oral inoculation of the young rabbit induces clear diarrhea as well as systemic infection. Using the young rabbit model, a *yst* mutant strain was defective in inducing diarrhea, weight loss, and death relative to wild-type *Y. enterocolitica* (Delor and

Cornelis, 1992). These data suggest that Yst may be a mediator of the diarrhea observed in infants infected with *Y. enterocolitica*. *yst* has been found in both pathogenic and non-pathogenic strains of *Y. enterocolitica*. Homologous DNA was also found in some strains of *Y. kristensenii* (Delor *et al*., 1990).

1.3.2 Mucoid *Yersinia* factor (Myf)

Yersinia species also possess a chromosomal gene which encodes a fibrillar protein in pathogenic serotypes of *Y. enterocolitica* known as Myf (Cornelis, 1994). Myf is composed of MyfA (a 21 kDa protein), MyfB (a chaperone) and MyfC (an outer-membrane protein) (Iriarte and Cornelis, 1995). As with other virulence factors, the fibrillar protein has homology to proteins with similar function in enterotoxigenic *E. coli*. These proteins are hypothesized to work in conjunction with Yst to aid in the colonization of mucous membranes. *Y. pseudotuberculosis* and *Y. pestis* have a protein with similar function of Myf, but is called pH6 antigen or pilus adhesion (PsaA), which may be involved in defence once the organisms are within the phagocyte (Lindler and Tall, 1993).

1.3.3 Invasin (Inv)

Invasin (Inv), an outer membrane protein, mediates attachment and entry into nonphagocytic cells (Isberg *et al*., 1987) and binds tightly to a family of α/β1- integrins, which are host molecules that bind to extracellular matrix proteins such as fibronectin on the basolateral surface of epithelial cells (Isberg and Leong, 1990). By binding tightly to integrins, invasion mediates bacterial uptake via a “zipper”-like mechanism, zippering the host cell membrane around the bacterium as it enters (Isberg and Leong, 1990). Invasin, by itself, is sufficient to mediate this uptake; invasin cloned into noninvasive *E. coli*, or purified invasin coupled to inert particles such as beads, mediates particle uptake. Host actin is needed for bacterial uptake since cytochalasins inhibit particle uptake (Finlay and Falkow, 1988). However, cytoskeletal rearrangements are not dramatic and disappear within a few minutes of bacterial entry (Young *et al*., 1992). Host signal transduction mechanisms appear to be necessary for invasin-mediated bacterial entry, since host tyrosine kinase inhibitors that block host cell signaling prevent bacterial uptake but adhere to cultured cells (Rosenshine *et al*., 1992).

1.3.4 Attachment invasion locus (Ail)

One of the chromosomal genes that is highly correlated with virulence is *ail* (Miller and Falkow, 1988). Ail is a 17 kDa outer membrane protein that is predicted to have eight membrane-spanning amphipathic β-strands and four extracellular loops (Miller *et al*., 1990; Beer and Miller, 1992). Ail was identified along with *inv* in a screen for *Y. enterocolitica* genes that could confer an invasive phenotype to *Escherichia coli. E. coli* carrying *ail* are highly invasive for some tissue culture cell lines (CHO cells) and moderately invasive for other tissue culture cells (HEp-2 cells). Ail mediates a high level of adherence to both CHO and HEp-2 cells. It was subsequently demonstrated that Ail also confers a high level of serum resistance to *E. coli* (Bliska and Falkow, 1992; Pierson and Falkow, 1993). Mutations in *ail* in *Y. enterocolitica* reduce the ability of *Y. enterocolitica* to invade tissue culture cells and significantly reduce their ability to survive the bactericidal effects of serum (Pierson and Falkow, 1993). Interestingly, *ail* sequences are only found in the pathogenic yersiniae (Miller *et al*., 1989). The *ail* gene of *Y. pseudotuberculosis* does not confer the attachment and invasion phenotype to *E. coli* but does confer high levels of resistance against serum complement, and an *ail* mutant of *Y. pseudotuberculosis* is serum sensitive (Yang *et al*., 1996).

1.3.5 *Yersinia* adhesin (YadA)

The adhesion protein YadA is encoded by the *yadA* gene located in the virulence plasmid of *Yersinia* (pYV) that is common to the pathogenic *Yersinia* species (Bukholm *et al*., 1990; Heesemann *et al*., 1984; Isberg, 1989). YadA is a virulence factor of *Y. enterocolitica*, however, YadA seems to be dispensable for the virulence of *Y. pseudotuberculosis*, and in wild-type *Y. pestis* the *yadA* gene has a frameshift mutation silencing the gene (Rosqvist *et al*., 1988; Skurnik and Wolf-Watz, 1989). Expression of the *Y. pseudotuberculosis* YadA in *Y. pestis* reduces its virulence (El Tahir and Skurnik, 2001). YadA is a homotrimer of ca. 45-kDa subunits that are anchored to the outer membrane via their C-termini, while their N-termini form a globular head on top of a stalk; the 'lollipop'-shaped YadA structure covers the entire bacterial surface giving it hydrophobic properties (Gripenberg-Lerche *et al*., 1995; Hoiczyk *et al*., 2000; Mack *et al*., 1994; Skurnik *et al*., 1984). The *yadA* gene expression is induced at 37°C by the temperature-dependent transcriptional activator LcrF. YadA is a multifaceted protein as revealed by its different biological properties. YadA+ bacteria bind to collagens, laminin, fibronectin, intestinal submucosa, mucus, and to hydrophobic surfaces like polystyrene. YadA+ bacteria autoagglutinate in stationary culture and also specifically

agglutinate guinea pig red blood cells. YadA is also a potent serum resistance factor as it inhibits the classical pathway of complement. As invasin, it mediates low rate invasion to tissue culture cells. In a rat model of reactive arthritis YadA and specifically YadA-mediated collagen binding is necessary for *Y. enterocolitica* to induce the disease (Balligand *et al*., 1985; Han and Miller, 1997; Tamm *et al*., 1993; Tertti *et al*., 1992).

1.3.6 The Yersiniabactin iron acquisition system

Iron acquisition is an essential requirement for all microorganisms except certain lactobacilli and *Borrelia burgdorferi* (Posey and Gherardini, 2000). One of the major differences between weakly and high pathogenic *Yersinia* lies in their ability to capture the iron molecules necessary for their systemic dissemination in the host. Several iron-uptake (siderophore) systems in enterobacteria are located on specific chromosomal regions designated pathogenicity islands. The high-pathogenicity island (HPI), initially found and characterized in *Yersinia* spp. may be considered an archetype of iron-uptake islands (Carniel, 2001). The siderophore synthesized specifically by highpathogenicity *Yersinia* is called yersiniabactin (Heesemann *et al*., 1993). This 482-Da molecule belongs to a small sub-group of phenolate siderophores and has an affinity for ferric iron ($K_D = 4 \times 10^{-36}$) much higher than for ferrous iron (Gehring *et al*., 1998). The 30- kb right-hand part of the HPI, termed the yersiniabactin locus, is highly conserved in the three pathogenic species *Y. pestis*, *Y. pseudotuberculosis* and *Y. enterocolitica*. This locus is composed of 11 genes organized in four operons. The exact function of all these genes is not yet entirely elucidated, but they can roughly be divided into three functional groups: yersiniabactin biosynthesis, transport into the bacterial cell (outer membrane receptor and transporters), and regulation. This pathogenicity island has never been detected in weakly pathogenic or avirulent strains of *Yersinia* (de Almeida *et al*., 1993).

1.3.7 *Yersinia* outer proteins (Yops)

The virulence plasmid, referred to as the pYV plasmid in *Y. enterocolitica*, encodes the Yop virulon – an integrated virulence apparatus (Cornelis *et al*., 1998; Iriarte *et al*., 1998). These proteins are involved in the resistance of *Yersinia* to phagocytosis by PMNs and macrophages, in inhibition of the PMN oxidative burst (Lian and Pai, 1985), in induction of programmed cell death in macrophages (Mills *et al*., 1997), and in inhibition of the cytokine release that is normally induced by *Yersinia* infection (Schulte *et al*., 1996), so limiting the host's inflammatory response to the infection. The Yop virulon includes a set of secreted

proteins (the Yop proteins), the Yop-specific chaperones (the Syc proteins), a Yop-dedicated secretion system (the Ysc system), and a regulatory network. Some of the secreted Yop proteins are required for the translocation across the eukaryotic cell membrane of other Yops that interfere with normal cellular processes in the cytosol of the eukaryotic cell.

The Ysc system belongs to the family of type III secretion systems, which are called contact-dependent because intimate contact between the eukaryotic cell and bacteria triggers secretion and allows delivery of bacterial proteins inside eukaryotic cells (Fallman *et al*., 1997). Homologous secretion systems are also present in a number of other bacteria, all of which interact with eukaryotic cells. These include animal pathogens – *Salmonella enterica*, *Shigella flexneri*, enteropathogenic and enterohaemorraghic *E. coli* (EPEC and EHEC), *Pseudomonas aeruginosa*, etc. (Cornelis *et al*., 1998; Alfano and Collmer, 1997; Hueck, 1998).

In vitro, Yop secretion only occurs at 37ºC in the absence of Ca^{2+}. This secretion correlates with growth arrest, a phenomenon known for a long time as "Ca^{2+} dependency" (Boyd and Cornelis, 2001, in "Principles of Bacterial Pathogenesis"). There seem to be two different regulatory networks. The first permits full expression of all the virulence plasmid-encoded virulence functions when the environment is ideal and the temperature reaches 37ºC, while the second only prevents Yop production in the presence of 2,5 m*M* Ca^{2+} ions (Cornelis *et al*., 1991; Rohde *et al*., 1994; Bolin *et al*., 1988; Forsberg and Wolf-Watz, 1988).

1.3.8 Lipopolysaccharide (LPS)

Lipopolysaccharide (LPS) is the major component of the outer membrane of Gram-negative bacteria. Although much attention has been given to the biological effects of its lipid A portion, a great body of evidence indicates that its O chain polysaccharide (O antigen) portion plays an important role in the bacterium–host interplay. Skurnik and colleagues demonstrated that the LPS O antigen is also essential for full virulence of *Y. enterocolitica* serotypes O:3 and O:8 (al Hendy *et al*., 1992; Zhang *et al*., 1997). A mutant lacking the O antigen is 100-fold less virulent than the wild type after oral infection. An interesting feature of *Y. enterocolitica* O antigen is that its expression is temperature regulated. The optimum expression occurs when bacteria are grown at room temperature (RT, 22–25ºC). However, when they are grown at 37ºC, the host temperature, only trace amounts of O antigen are produced (al Hendy *et al*., 1991; Brubaker, 1991; Bengoechea *et al*., 2002). Significantly, most, if not all, *Yersinia* virulence factors are also regulated by temperature (Straley and

Perry, 1995). Recent work by Bengoechea and colleagues showed that the *Y. enterocolitica* O:8 LPS O antigen mutants of three different O antigen phenotypes (expressing no O antigen, one O unit or randomly distributed O antigen chain lengths) are attenuated in virulence regardless of the infection route used (Bengoechea *et al*., 2004). The results indicated that for the full virulence of *Y. enterocolitica* O:8 not only the presence of the O antigen but also the proper distribution of the O antigen chain lengths is required. They also demonstrated that the function or expression of other virulence factors of *Y. enterocolitica* O:8 require the presence of the O antigen.

2. Suppressive subtractive hybridization as a tool in the elucidation of the genetic variability among *Yersinia* strains

The presence of different sequences and virulence factors in similar bacterial pathogens suggests that bacterial diversity may commonly involve horizontal transfer, loss and acquisition of important blocks of chromosomal or plasmid DNA encoding a series of related gene products. For example, high pathogenic *Y. enterocolitica* BG 1B strains have a chromosomal "pathogenicity island" and a pYV encoded Yop virulon, which are absent in non pathogenic *Y. enterocolitica* BG 1A isolates (Carniel, 2001). On the other hand, even different strains belonging to the avirulent group, such as ST O:5 or ST O:6,30, BG 1A may cause an infection. Thus, serotype-, biogroup- or even strain-specific markers might be responsible for the differences in clinical manifestations of yersiniae. To map out such lineage-specific markers, we have applied the suppression subtractive hybridization (SSH) method (Diatchenko *et al*., 1996; Lisitsyn *et al*., 1993; Lisitsyn and Wigler, 1995). This method was already successfully used for the identification of genomic differences between *Y. enterocolitica* WA-C BG 1B, ST O:8 and *Y. enterocolitica* NF-O biogroup 1A, ST O:5 (Iwobi *et al*., 2002). The *Y. enterocolitica* species represent a highly heterogeneous group of bacteria ranging from the non pathogenic BG 1A strains to the high pathogenic BG 1B and low pathogenic isolates of BG 2-5 (Table 3). Although closely related, the BG 1B and BG 2-5 organisms differ significantly with respect to pathogenicity.

3. Diagnostics of *Yersinia*

The gold standart for the final biochemical identification of a *Yersinia* isolate is still classical tube testing (Aleksic and Bockemuhl, 1990). Commercially available identification systems such as the API 20E, API rapid 32 IDE, the GNI card of the VITEK system (all bioMerieux)

or the Micronaute E (Merlin Diagnostika) lack sensitivity and specificity at the species level (Neubauer *et al*., 1998; Linde *et al*., 1999). The definite identification of *Y. enterocolitica* isolates can also be achieved by sequencing the 16S rRNA gene (Neubauer *et al*., 2000). A commercially available PCR kit and a dip stick probe assay for the identification of *Y. enterocolitica* isolates based on 16S rRNA gene sequences proved to be specific either for "European" or "American" isolates (Neubauer *et al*., 1999).

For the assessment of the presumptive pathogenicity of an *Y. enterocolitica* isolate a variety of empiric virulence markers are used (Aleksic and Bockemuhl, 1990). These assays can easily be replaced by PCR assays targeting plasmid genes (Neubauer *et al*., 2000; Neubauer *et al*., 2000). The evaluation of molecular systems in terms of routine diagnostic procedures, however, is still missing.

4. Aims of this research study

Depending on the pathogenicity level, weakly and highly pathogenic *Yersinia* strains use different strategies for their interactions with a host. Discovering chromosomal and plasmid determinants, which are responsible for the different strategies utilized by weakly and highly pathogenic yersiniae will allow us to trace the evolution of pathogenicity in yersiniae. Because of problems with typing of *Y. enterocolitica* strains as well as their serological discrimination from *Brucella* species, pathotype-specific genomic determinants will be good epidemiological markers.

In this work, we made an attempt to recognize genomic features of the weakly and highly pathogenic yersiniae, as well as to fish out signature sequences that will be suitable for epidemiological studies.

In our study, a mouse lethal *Y. enterocolitica* BG 1B, ST O:8, which presents the classic "American" highly pathogenic group of strains, was used both as a tester and as a driver in SSH with a mouse non virulent *Y. enterocolitica* BG 4, ST O:3 strain, which is a typical "non-American" isolate and mainly distributed in Germany, to uncover gene acquisitions and losses in both strains of the selected pair. We have also characterized the gene acquisitions of the representative Rha-positive, human avirulent *Y. pestis* strain G8786 isolated in Georgia in the high mountainous Caucasian locus from *Microtus arvalis* in 1987, by means of SSH with another human virulent Antiqua strain Yokohama.

This research study had two aims:

I. Uncovering gene acquisitions and losses in chromosomes of the two *Y. enterocolitica* isolates, namely *Y. enterocolitica* subsp. *palearctica* BG 4 ST O:3 (weakly pathogenic Y-108C strain) and *Y. enterocolitica* subsp. *enterocolitica* BG 1B ST O:8 (highly pathogenic WA-C strain) by SSH.

II. Characterization of gene acquisitions of the human avirulent *Y. pestis* strain G8786 compared with another Antiqua strain Yokohama by SSH.

This differential analysis approach should lead to the following goals:

1) Isolation and characterization of novel DNA sequences unique to the highly and weakly pathogenic *Y. enterocolitica* BG 4 and BG 1B strains.

2) Finding a novel virulence factors and epidemiological markers specific to the weakly pathogenic *Y. enterocolitica* BG 4 strains

3) Isolation and characterization of novel DNA sequences unique to the human avirulent *Y. pestis* strain G8786

B. MATERIALS AND METHODS

1. Material

1.1 Equipment

Centrifuge	Sigma, Deisenhofen 1K1S, Table-centrifuge 3K30 with Rotor Nos. 12156 and 19776
Electrophoresis apparatus for SDS-PAGE	Bio-Rad, München Mini-Protean -II Cell and Western Blot Apparatus
Electroporation-apparatus	Bio-Rad, München, Gene Pulser, II Pulse Controller II
French Press	French Pressure Cell 40K SLM Aminco
Hybridization oven	Personal Hyb. Stratagene, Amsterdam
Incubator	Heraeus, Hanau Typ B20
PCR-Cycler	PE Applied Biosystems, Weiterstadt Gene Amp 2400
pH Meter	Mettler, Toledo 320 pH Meter
Photometer	Pharmacia, Biotech Ultrospec 2000
Pipettes	Eppendorf, Hamburg Research P10-P1000
Sequencer	PE Applied Biosystems, Weiterstadt ABI 377 DNA Sequencer
Shaking incubator	Braun, Melsungen Certomat BS-1
Sterile bank	Heraeus, Hanau Herasafe HS12
Transilluminator	Heralab, Wiesloch, UVT-20M/W
Video-equipment	Sigma, Deisenhofen, EASY (Enhanced Analysis System)
Vacuum blot	Pharmacia,-LKB, Uppsala, LKB 2016 Vacu Gene^{R}-Chamber
Scale	Sartorius, Göttingen Model R 160P and Pt 1200

1.2 Other materials

Plastic and related articles were purchased from the following firms: Nunc, Roskilde, DK; Sartorius, Göttingen; Falco/Becton Dickinson, Heidelberg; B. Braun, Melsungen; Eppendorf, Hamburg; Greiner, Nürtingen and Schleicher & Schüll, Dassel.Nylon membranes (Zeta Probe

GT) were purchased from Biorad and Nitrocellulose membranes (Whatman-paper) from Schleicher & Schüll.

1.3 Chemicals and Enzymes

All chemicals and antibiotics were supplied by Merck (Darmstadt), Biochrom (Berlin), Roche (Mannheim) and Sigma (Deisenhofen). Media plates were supplied by Difco (Detroit, Michigan, USA) and enzymes were obtained from MBI Fermentas (St. Leon-Roth), Roche (Mannheim), and Gibco (Eggenstein).

2. Bacteria, Plasmids and Primers

2.1 Table 4. Bacterial strains and plasmids

Strain / plasmid	Relevant Characteristics	References or source
Srains		
E. coli		
JM109	{ *recA1 endA1 gyrA96 thi hsdR17*(r_k-, m_k+) *supE44 relA1* Δ(*lac-proAB*) [F' *traD36 proAB lacI*^{q}ZΔM15]}	(Yanisch-Perron *et al.*, 1985)
JM109 Nal^r	The spontaneous Nal^R mutant of *E. coli* strain JM109	This research
XL1 Blue MR	(*mcrA) 183 (mcrCB-hsdSMR-mrr) 173 endAI supE44 thi-1 recA1 gyrA96 relA1lac*^{c}	Stratagene
BL21	F$^-$ *ompT hsdS*(r_B^- m_B^-) *gal*	Stratagene
Y. enterocolitica		
Y-108-P	Clinical isolate, ST O:3 BG 4	(Heesemann *et al.*, 1984)
Y-108-C	Plasmidless variant of Y-108P	This research
Y11	Clinical isolate, ST O:3 BG 4	SaBw strain collection, Dr. Neubauer
5307	Clinical isolate, ST O:3 BG 4	SaBw strain collection, Dr. Neubauer
5783	Clinical isolate, ST O:3 BG 4	SaBw strain collection, Dr. Neubauer
Y486	Animal isolate, ST O:3 BG4	SaBw strain collection, Dr. Neubauer
Y626 WS 15/93	Animal isolate, ST O:3 BG4	SaBw strain collection, Dr. Neubauer
Y633 (WE50/01)	Clinical isolate, ST O:3 BG 4	SaBw strain collection, Dr. Neubauer
Y641	Animal isolate, ST O:3 BG4	SaBw strain collection, Dr. Neubauer
560 (SW13123)	Animal isolate, ST O:3 BG4	(Saken *et al.*, 1994)
559 (M388)	Animal isolate, ST O:3 BG4	(Saken *et al.*, 1994)
556 (8265)	Clinical isolate, ST O:3 BG 4	(Saken *et al.*, 1994)

531 (SW13711)	Animal isolate, ST O:3 BG4	(Saken *et al*., 1994)
554 (29C-43a)	Clinical isolate, ST O:3 BG 4	(Saken *et al*., 1994)
555 (MCH697)	Clinical isolate, ST O:3 BG 4	(Saken *et al*., 1994)
Y485	Animal isolate, ST O:3 BG4	SaBw strain collection, Dr. Neubauer
628 W 498	Animal isolate, ST O:3 BG4	SaBw strain collection, Dr. Neubauer
642	ST O:3 BG4	SaBw strain collection, Dr. Neubauer
Y631	Animal isolate, ST O:3 BG4	SaBw strain collection, Dr. Neubauer
S-2840	Clinical isolate, ST O:3 BG 4	MvP strain collection
2118-Y	Clinical isolate, ST O:3 BG 4	MvP strain collection
221 Erg.Nod	Clinical isolate, ST O:3 BG 4	MvP strain collection
Y-1088	Clinical isolate, ST O:3 BG 4	MvP strain collection
66 Stuhl	Clinical isolate, ST O:3 BG 4	MvP strain collection
56 LK	Clinical isolate, ST O:3 BG 4	MvP strain collection
62 Stuhl	Clinical isolate, ST O:3 BG 4	MvP strain collection
59 LK	Clinical isolate, ST O:3 BG 4	MvP strain collection
4147 pYVII Ar+	Clinical isolate, ST O:3 BG 4	MvP strain collection
7347 pYVI Ar+	Clinical isolate, ST O:3 BG 4	MvP strain collection
Y745 IP24231	ST O:3 BG 3, Man, Japan	SESAHS strain collection, Dr. Pham
Y746 IP24232	ST O:3 BG 3, Man, Japan	SESAHS strain collection, Dr. Pham
Y747 IP134	ST O:3 BG 4, Man, Sweden	SESAHS strain collection, Dr. Pham
Y748 IP21981	ST O:3 BG 4, Man, France	SESAHS strain collection, Dr. Pham
Y749 IP1601	ST O:3 BG 4, Man, Japan	SESAHS strain collection, Dr. Pham
Y750 IP19718	ST O:3 BG 4, Man, China	SESAHS strain collection, Dr. Pham
Y751 IP23222	ST O:3 BG 4, Man, UK	SESAHS strain collection, Dr. Pham
Y752 IP23357	ST O:3 BG 4, Man, Brazil	SESAHS strain collection, Dr. Pham
Y753 IP24309	ST O:3 BG 4, Man, N. Caledonia	SESAHS strain collection, Dr. Pham
Y754 IP24310	ST O:3 BG 4, Man, N. Caledonia	SESAHS strain collection, Dr. Pham
Y755 IP7032	ST O:3 BG 4, Man, South Africa	SESAHS strain collection, Dr. Pham
Y756 IP7036	ST O:3 BG 4, Man, South Africa	SESAHS strain collection, Dr. Pham
Y757 IP3692	ST O:3 BG 4, Man, Hungary	SESAHS strain collection, Dr. Pham
Y758 IP3704	ST O:3 BG 4, Man, Hungary	SESAHS strain collection, Dr. Pham
Y759 IP4115	ST O:3 BG 4, Man, Canada	SESAHS strain collection, Dr. Pham
Y763 IP4124	ST O:3 BG 4, Man, Canada	SESAHS strain collection, Dr. Pham
Y764 IP4125	ST O:3 BG 4, Man, Canada	SESAHS strain collection, Dr. Pham
Y765 IP22274	ST O:3 BG 4, Man, Australia	SESAHS strain collection, Dr. Pham
Y766 IP22275	ST O:3 BG 4, Man, Australia	SESAHS strain collection, Dr. Pham
Y767 IP22276	ST O:3 BG 4, Man, Australia	SESAHS strain collection, Dr. Pham
Y768 IP25728	ST O:3 BG 4, Man, Australia	SESAHS strain collection, Dr. Pham

Y769 IP23227	ST O:3 BG 4, Man, New Zealand	SESAHS strain collection, Dr. Pham
Y770 IP23228	ST O:3 BG 4, Man, New Zealand	SESAHS strain collection, Dr. Pham
Y771 IP23230	ST O:3 BG 4, Man, New Zealand	SESAHS strain collection, Dr. Pham
910/98	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
562/03	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
465/00	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
128/99	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
627/99	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
120/98	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
243/96	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
504/97	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
99/96	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
457/98	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
567/98	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
265/01	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
274/99	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
188/01	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
254/97	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
353/96	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
332/98	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr. Gierczynski
184/97	Clinical isolate, ST O:3 BG 4, Poland	PZH strain collection, Dr.

		Gierczynski
96C (Y-96)	Clinical isolate, ST O:9 BG 2	(Saken *et al*., 1994)
564 (7Oulua)	Clinical isolate, ST O:9 BG 2	(Saken *et al*., 1994)
563 (YE099)	Clinical isolate, ST O:9 BG 2	(Saken *et al*., 1994)
Y738 IP22393	ST O:9 BG 2, France	SESAHS strain collection, Dr. Pham
586 (H567/90)	Clinical isolate, ST O:5,27 BG 2	(Saken *et al*., 1994)
568 (YE873)	Animal isolate, ST O:5,27 BG 2	(Saken *et al*., 1994)
534 (D113)	Animal isolate, ST O:5,27 BG 2	(Saken *et al*., 1994)
Y740 IP199	ST O:5,27 BG 2, Man, USA	SESAHS strain collection, Dr. Pham
Y741 IP885	ST O:5,27 BG 2, Man, UK	SESAHS strain collection, Dr. Pham
Y743 IP22460	ST O:5,27 BG 2, Man, Australia	SESAHS strain collection, Dr. Pham
JD E029	Clinical isolate, ST O:1	(Schiemann, 1982)
gk1142	Hare, ST O:2 BG 5	(Kapperud *et al*., 1985)
JD E766	Clinical isolate, ST O:1,2,3	(Schiemann, 1982)
Y772 IP1	ST O:2a,2b,3 BG 5, Hare, France	SESAHS strain collection, Dr. Pham
Y773 IP178	ST O:2a,2b,3 BG 5, Hare, France	SESAHS strain collection, Dr. Pham
438/80	ST O:6,31	(Skurnik, 1984)
1309/80	ST O:6,31	(Skurnik, 1984)
IP 2222	Non-clinical isolate, ST O:36, BG 1A	(Grant *et al*., 1999)
gc 3973-76	Clinical isolate, ST O:4	(Kay *et al*., 1983)
JD E701	Clinical isolate, ST O:4,32 BG 2	(Schiemann, 1982)
NF-O	Clinical isolate, ST O:5 BG 1A	(Saken *et al*., 1994)
Y755 IP124	ST O:5 BG 1A, Pony, France	SESAHS strain collection, Dr. Pham
Y774 IP102	ST O:6,30-6,61, Man, Denmark	SESAHS strain collection, Dr. Pham
189/80	ST O:6,30 BG 1A	(Skurnik, 1984)
6737/80	ST O:6,30 BG 1A	(Skurnik, 1984)
350	ST O:7,8	MvP strain collection
329/82	ST O13:b,e,f,i	MvP strain collection
3039/85	ST 020:b,e,f,i	MvP strain collection
209/84	ST 021:b,e,f,I	MvP strain collection
478/91	ST O:41,43	MvP strain collection
4643/93	ST O:48	MvP strain collection
WA-C	Clinical isolate, ST O:8 BG 1B	(Heesemann *et al*., 1984)
589 (900/90)	Clinical isolate, ST O:8 BG 1B	(Saken *et al*., 1994)
575 (893/87)	Clinical isolate, ST O:8 BG 1B	(Saken *et al*., 1994)
Y736 IP636	ST O:8 BG 1B, USA	SESAHS strain collection, Dr. Pham
Y. pseudotuberculosis		
PB1	ST O:1	Michigan State University, USA, Dr.

		Brubaker
346	Clinical isolate, ST O:3	MvP strain collection
Y. pestis		
735	Rattus norvegicus	MvP strain collection
5764	Rattus norvegicus	MvP strain collection
952	Rattus norvegicus	MvP strain collection
G8786	Vole	(Golubov *et al*., 2004)
EV	Clinical isolate	MvP strain collection
Yokohama	Clinical isolate	MvP strain collection
Other yersiniae		
Y. intermedia H9-36/83	ST O:17	MvP strain collection
Y. frederiksenii H56-36/81	ST O:60	MvP strain collection
Y. kristensenii H25-36/84	ST O:50	MvP strain collection
Y. rohdei H274-36/78	ST O:76	MvP strain collection
Y. bercovieri H632-36/85	ST O:16	MvP strain collection
Y. mollaretii H279-36/86	ST O:59	MvP strain collection
Y. ruckeri 529-36/85	unknown	MvP strain collection
Y. aldovae H344-36/91	ST O:NT	MvP strain collection
Plasmids		
SuperCos1	Cloning vector Ap^r Neo^r	Stratagene
pSC1275D	SuperCos1 carrying a proximal part of the *rtx* operon	This research
pSC1012B	SuperCos1 carrying a distal part of the *rtx* operon	This research
pMOS*Blue*	Cloning vector Ap^r	Amersham
pET100/D-TOPO	Expression vector which allows expression of recombinant protein with an N-terminal 6xHis tag	Invitrogen
pRTX2100	pET100D/TOPO carrying 2100 bp of *rtxA* gene	This research
pG8786	pFra plasmid from *Y. pestis* G8786 strain	(Golubov *et al*., 2004)
pG8786-Cm^r	pG8786 carrying Cm^r gene	(Golubov *et al*., 2004)
RP4	the conjugative plasmid	(Kim *et al*., 1993)
pKD46	plasmid carrying red system genes	(Murphy, 1998)
pKD3	Plasmid carrying Cm^r gene was used as	(Murphy, 1998)

	a source of Cm^r cassete	

2.2 Table 5. List of primers: SSH (seq. primers), IS*1331*, *rtx*, pG8786

Primer	5'-…..-3' sequence	Description
Primers for SSH project		
T7 (forward)	TAATACGACTCACTATAGGGA	amplifies the insert within the MCS of pMOS*Blue* vector
U19 (reverse)	GTTTTCCCAGTCACGACGT	amplifies the insert within the MCS of pMOS*Blue* vector
Primers for IS*1331* project		
istBL.for	TAAGAAAATGCCACTTGTGAGGGG	sequencing primer
istAL.for	CGTTAATCGTTGTTTGGCGGG	sequencing primer
istAM.for	TCCGTAATGTTCTCACTTGACGCTC	sequencing primer
istAL2.for	GGGCAAATACCCCAAATCATC	sequencing primer
ist.for	GGTTTCATCAGTTAGGCGTTG	sequencing primer
ist2.for	GATTAATACGGTAGCCAATACTCG	sequencing primer
istAL3.for	ATGCTGGGTTTATTTAATGGGC	sequencing primer
gpP.for	GGCATCAATTCCGATATATTCCACG	sequencing primer
orf19.rev	CGGAGGTTGAGAATCTATTCCAGAA GAG	sequencing primer
O1R.rev	TGGAGTAGCTCAGTAAAGGTGG	sequencing primer
orf155m.rev	AAGATTGGCCCCAACAGAGTC	sequencing primer
O2R.rev	GTGCAAAAGTAGACCAAGGGATC	sequencing primer
istL.rev	GAATGGGTAAAACTATTTGCCGATG	sequencing primer
fic.for	TATGCTGAAGGCTGTGGCTGTCA	sequencing primer
istBN2.for	TTATGCTGTGGGAACCTGGCTTG	sequencing primer
istBL.rev	CCACCAATGGGTTTAAAGTTCGC	sequencing primer
istAR.rev	CAAGTGCCGTCTGCCTGTTGA	sequencing primer
istAM.rev	AAGAACAATATCAACCCGCGCC	sequencing primer
istAR1.rev	GCTCTCGCTGTGAAATCCGGC	sequencing primer
ist.rev	AAATTACCCGCCAAACAACG	sequencing primer
ist2.rev	CACAGCATAAAACACTGTCTGAC	sequencing primer
istAR2.rev	CGTCGCGCTGCTTTTGCTGT	sequencing primer
intU.rev	TTATATTATCGGGGTGCATGGCAC	sequencing primer

O1L.for	ATTGGTAATACGCAGAGTTTCCTTC	sequencing primer
orf181m.for	AATGTGCACAGAGAAGATAATCG	sequencing primer
O2L.for	GCCCTAGTCGCTTTTTACCTAATG	sequencing primer
istR.for	CGGTATGACGAGATATTTGTAACTG	sequencing primer
STY4827.for	ATGAGCCAAACAACCCCACACCC	sequencing primer
istBN2.rev	CATCATTGTGACATTATTGAAACAG GCA	sequencing primer
istAN1.rev	TCAATGCCATCAGATCATCCAAACA	sequencing primer
Primers for pG8786 project		
YopP8	GAGACCAGTTCTTTAATCAG	amplifies part of *yopP*
YopP9	GCCAGTGCCAAACTAAAAAT	amplifies part of *yopP*
cafD.for	CTGACAAATTTATGTGAAGATCAAT GTTAGGAACTAATGCAGAAAGCCA CGGTGTAGGCTGGAGCTGCTTC	generates PCR fragment for ET mutagenesis
cafD.rev	AACCCCGGGGTGAGGGCAAAGGCT GCTTTGTTGAAGTTGCATGGATGAT GGCATATGAATATCCTCCTTAG	generates PCR fragment for ET mutagenesis
cafD1.for	GGGGATGACGTCGTCTTGGCTAC	amplifies part of *cafD*
cafD1.rev	TCCACTCACTGAGTGAAGCCCTTTT AA	amplifies part of *cafD*
Primers for *rtx* project		
delA.for	ATGCCCGTGGGTGGGAATGATGATA TTTTCGTCGGATCCATTGCTGCAGT GTAGGCTGGAGCTGCTTC	generates PCR fragment for ET mutagenesis
delA.rev	CTTAGCTTTCTCCACATTCTTTCTTC CCTCCACTAAAGCTGATTCAGACAT ATGAATATCCTCCTTAG	generates PCR fragment for ET mutagenesis
alf10.for	ATACCATTCAAGTCGGAATGATGAG	sequencing primer
alf11.for	AACCTCGACGAAATCAAGTCAACTG	sequencing primer
alf12.for	GGAAGATCTGCTTAATATGGAGCAG	sequencing primer
alf13.for	CGGAAGATCTGCTTAATATGGAGCA G	sequencing primer
alf14.for	GAAGGAGATGATGAACTTATCGCTG	sequencing primer
alf141.for	TCTCTAGCGGGGCTGAAAAAG	sequencing primer
alf15.for	CAGGTTTTGAGATAAATTATGCCCG	sequencing primer
rtxA_1.rev	TTACACCCCCCCCATCAGAGCATAA GTG	generates PCR fragment with the first part of *rtxA* for

		subcloning into the expression vector pET100/D-TOPO
alf16.for	ATATTCATCCATTATCTACGGCGG	sequencing primer
alf17.for	GGAGTAATGTCCACACCGTATAAG	sequencing primer
alf18.for	TTAGGTGAATTGCCCCTGAGTG	sequencing primer

All the primers used in this work were supplied by Metabion (Martinsried). They were supplied in either a 100 pmol/µl solution or lyophilized. Lyophilized primers were dissolved in distilled, sterile water to a 100 pmol/µl end concentration. Table 2 gives a list of the primers used in this work.

3. Culture media, Antibiotics, Strain Cultivation and Storage

3.1 Culture media

Sterilization of liquid media was by autoclaving (121 °C at 1 bar for 20 min). For solid agar, 15 g agar per liter of liquid media was used.

Luria-Bertani (LB) Medium	Trypton	10 g
	Yeast powder extract	5 g
	NaCl	5 g
	H_2O_{dest} to	1 L
	pH set to 7.4 - 7.6 with NaOH	
Minimal medium (M9)	Na_2HPO_4	6 g
	KH_2PO_4	3 g
	NaCl	0.5 g
	NH_4Cl	1 g
	Water to	1 L
	pH adjusted to 7.4, autoclaved and cooled	
Further media components	1M $MgSO_4$	
	20% Glucose	
	1M $CaCl_2$	

0.5M EGTA

3.2 Antibiotics

Name and concentration of antibiotics employed in this research are listed in table 3. Sterilization of all antibiotics was by filtration with 0.22 µm filters.

Table 6. List of Antibiotics

Antibiotic	Abbreviation	Dissolved in	End-concentration (µg/ml)
Ampicillin	Amp	H_2O_{dest}	100 (*E. coli*)
			400 (*Yersinia*)
Kanamycin	Km	H_2O_{dest}	50
Nalidixic acid	Nal	0.5 N NaOH	100
Chloramphenicol	Cm	70 % C_2H_5OH	30
Streptomycin	Sm	H_2O_{dest}	100
Tetracycline	Tet	70 % C_2H_5OH	15

3.3 Cultivation and long term storage of bacteria

Bacteria were cultivated either on agar plates or in liquid medium by incubation on a shaker as follows:

- *Yersinia*: 200 rpm at 27°C
- *E. coli*: 200 rpm at 37°C. For long term storage, bacteria were suspended in LB-Medium fortified with 10% Glycerol and frozen at -80°C. Table 4 presents a summary of the strains and plasmids used in this study.

4. Molecular genetic methods

4.1 Isolation of Chromosomal DNA

High quality chromosomal DNA was isolated according to the method described in Current Protocols (Ausubel *et al*., 2000) as follows. Bacteria from a saturated liquid culture were lysed with SDS and proteinase K. Cell wall debris, polysaccharides, and remaining proteins were then removed by selective precipitation with CTAB (hexadecyl trimethyl ammonium bromide), and high molecular weight DNA was recovered from the resulting supernatant by isopropanol precipitation.

Reagents

Tris-EDTA Buffer 2M Tris.HCl, pH 8

	0.25M EDTA, pH 8
CTAB/NaCl solution	10% CTAB (hexadecyltrimethyl ammonium bromide) 0.7M NaCl (4.1g / 100 ml) Dissolve in 80 ml distilled water, heat to 65°C and adjust volume to 100 ml

Procedure

50 ml LB medium (supplemented with antibiotics when appropriate) was inoculated with the bacteria of interest and grown overnight on a shaker at the 27°C/37°C. The cells were then pelleted at 4,000 x g for 5 min and gently resuspended in 9.5 ml TE buffer. 0.5 ml of 10% SDS was added and 50µl proteinase K (20 mg/ml in H2O), followed by gentle mixing and incubation for 1 hr at 37°C. 1.8 ml 5M NaCl was added and mixed thoroughly. 1.5 ml CTAB/NaCl was then added, the suspension mixed and then incubated at 65°C for 20 min. 1 volume (13.5 ml) of freshly prepared chloroform/isoamylalcohol mixture (at a ratio of 24:1 respectively) was then added, and the suspension was centrifuged at 6,000 x g for 10 min at RT. The viscous upper phase was then drawn out with a tipless 5 ml pipette, mixed with 1 volume of Phenol/Chloroform/Isoamylalcohol (25:24:1) and subjected to centrifugation for at least 10 min at 6,000 x g at RT. This extraction step with phenol was repeated thrice, after which the last phase was collected in a fresh tube and treated with 0.6 volume isopropanol to precipitate the DNA. The solution was mixed carefully till the DNA self-precipitated. The DNA was then spooled out of the mix with a pipette tip and washed twice in 1 ml of 70% ethanol. The DNA was allowed to air-dried and finally dissolved overnight in high purity water.

4.2 Isolation of plasmid DNA

4.2.1 Plasmid isolation with QIAprep Spin Miniprep kit (Qiagen)

The QIAprep Spin Miniprep kit was routinely used for small scale isolation of plasmid DNA (up to 20 µg). The principle behind it is based on alkaline lysis, coupled with anion-exchange chromatography. The isolation procedure was as recommended by the kit’s manufacturer.

4.2.2 Plasmid isolation with Nucleobond AX100 Kit (Machery-Nagel)

The Nucleobond AX100 Kit was used for the isolation of high quality DNA in high concentration (up to 100 µg). The principle of DNA isolation is also based on alkaline lysis of cells, followed by purification of nucleic acids on the basis of anion-exchange chromatography. The isolation procedure was as recommended by the kit's manufacturer.

4.3 Purification DNA and determination of DNA concentration and purity

4.3.1 Phenol extraction and ethanol precipitation of DNA

Phenol extraction was carried out to remove contaminating proteins from a DNA preparation.

Procedure

- The DNA solution was mixed with an equal volume of TE - saturated phenol/chloroform/isoamyl alcohol (25:24:1) in a microcentrifuge tube and the mixture vortexed for 30 sec.
- The mixture was centrifuged at 14,000 rpm for 5 min at RT to separate the sample into phases.
- The upper aqueous layer was then removed into a clean tube, carefully avoiding denatured proteins found at the aqueous / phenol interface. This upper phase was then mixed with an equal volume of the phenol / chloroform / isoamyl alcohol solution mentioned above, the mixture vortexed and centrifuged (14,000 rpm for 5 min). This step was repeated 2-3 times, and the DNA precipitated from the upper aqueous phase through ethanol precipitation.

Ethanol precipitation

This was carried out to remove contaminating salts from a DNA preparation or to concentrate a DNA preparation.

Procedure

- The DNA solution was mixed with 1/10 volume of 3 M sodium acetate and 3 volumes of ethanol.
- The mixture was incubated at -20°C for 30 min.
- The mixture was centrifuged at 14,000 rpm for 15 min at 4°C.
- The supernatant was removed and the DNA pellet was washed with 70% ethanol and centrifuged at 14,000 rpm for 5 min at 4°C.
- The pellet was air-dried and the DNA resuspended in water and stored at -20°C.

4.3.2 Determination of DNA concentration and purity

Nucleic acids have a maximum absorption at 260nm wavelength. The isolated DNA was diluted with distilled water (1:100) and the absorbance at 260nm against H_2O_{bidest} measured spectrophotometrically. The calculation of the DNA concentration was based on the following formula:

1 A260 = 50 µg/ml for dsDNA

1 A260 = 33 µg/ml for ssDNA

For determination of DNA purity, the A260/280 coefficient was photometrically determined. An A260/280 < 1.8 indicated contamination of the DNA preparation with protein or aromatic substances such as phenol, while an A260/280 > 2.0 indicated possible contamination with RNA (LAB FAQs, Roche).

4.4 Polymerase Chain Reaction (Saiki *et al*., 1988)

The polymerase chain reaction (PCR) permits the selective in vitro amplification of a particular DNA region by mimicking the phenomenon of in vivo DNA replication. Typically, three steps are involved in a standard PCR recation: denaturation, which achieves the dissociation of the template DNA molecules into single strands; annealing, which allows single stranded primers to bind to complementary sites on the template DNA; and lastly elongation which allows for extension of the DNA strands, due to the effect of the thermostable DNA polymerase. As template DNA, either plasmid, cosmid or chromosomal DNA was utilized at a diluted concentration, or cooked cells were employed. Where cooked cells were used as source of template DNA, the procedure was as follows:

- A bacterial colony was isolated, suspended in 70 µl H2Obidest, cooked at 95 °C for 10 min and centrifuged (12,000 rpm for 5 min).
- The supernatant containing released DNA template was then utilized in the PCR reaction.

For a typical 50 µl reaction volume, the following components were pipetted into a PCR test-tube:

Reaction components	DNA	5 µl
	Primer 1 (100 pmoles)	0.5 µl
	Primer 2 (100 pmoles)	0.5 µl
	dATP, dCTP, dGTP, dTTP	5 µl
	10 x Taq-Reaction buffer	5 µl

	Taq-polymerase (5 U/µl)	0.5 µl
	H_2O	ad to 50 µl
Cycling parameters	Denaturation 94°C	3 min
	Denaturation* 94°C	30 sec
	Annealing* x°C	30 sec
	Elongation* 72°C	y min
	Final extension 72°C	3 min

* 30 - 35 cycles

x: Annealing temperature dependent on the Tm (melting temperature) of primers

y: Elongation is typically 1 min pro kb of amplified DNA

A negative control with water as template DNA was always included in the reactions and 5 µl of the finished PCR product was checked on an agarose gel before purification with the Qiagen PCR purification kit.

4.5 Agarose gel electrophoresis

The agarose gel was prepared by mixing an appropriate proportion of agarose (to a final concentration of 0.7 - 2% depending on the MW of the sample DNA) with 1 x TAE buffer, the mixture cooked and after cooling poured into precast agarose gel chambers. The DNA was then mixed with loading buffer, loaded onto spurs on the gel and electrophoretically separated by voltage application utilizing the 1 x TAE solution as the running buffer. Following the electrophoretic run, gels were stained in ethidium bromide solution and the DNA visualized under ultraviolet radiation.

Solutions:

1x TAE buffer	40mM Tris / HCl, pH = 8.2 20mM Acetic acid 2mM EDTA, pH 7.6
10 x Loading buffer for agarose gels	0.25% (w/v) Bromophenol blue 10% (v/v) Glycerol
Agarose gel	0.8% - 2% Agarose in 1 x TAE
Ethidium bromide staining solution:	1 µg Ethidium bromide pro ml H_2O

4.6 Enzymatic modification of DNA

4.6.1 Restriction digestion of DNA

Chromosomal or plasmid DNA samples were routinely subjected to restriction digestions. For a restriction endonuclease reaction, the following components were mixed together and incubated at the appropriate temperature (usually 37°C for most enzymes):

DNA	x µl
10 x Reaction buffer[x]	1 µl
Enzyme	2 - 3 units / µg DNA
H_2O	to 10 µl*

* For higher DNA concentrations, the reaction and volume were scaled up linearly.

x: Choice of reaction buffer depended on the type of enzyme employed.

Because all reaction enzymes are supplied in 50% glycerol, which can exert an inhibitory effect on digestion efficiency, care was taken that the glycerol concentration did not exceed 5% final digestion volume. Enzyme inactivation was either through heat treatment at 65°C for 20 min (Lab FAQs, Roche).

4.6.2 Dephosphorylation of DNA

This procedure removes the phosphate ends arising after digestion of a vector/plasmid DNA with restriction endonucleases, thus preventing dimerization or self-religation of vector or plasmid DNA. The vector DNA is then free to ligate with an insert DNA of choice. Shrimp alkaline phosphatase (SAP from Roche, Mannheim) was employed and the reaction proceeded at 37°C for 30 minutes, followed by heat inactivation at 70°C for 20 min.

4.6.3 Ligation of DNA molecules

Ligation of linear DNA molecules was with the enzyme T4 DNA ligase (Gibco, Eggenstein). Typically, a 1:2 vector to insert ratio was utilized for all ligations and the reaction proceeded at 16°C overnight.

4.7 DNA sequencing

DNA Sequencing was done by the dideoxy-chain terminating method on an automated ABI Prism DNA Sequencer. The ensuing chromatograms were processed with Chromas software and BLASTN and BLASTX programs provided by NCBI (National Center for Biotechnology Information) and TIGR (The Institute for Genomic Research), and also the *Y. pestis* and *Y. enterocolitica* gene banks from Sanger Center were employed for in-depth homology searches.

4.8 RNA analysis

4.8.1 RNA Isolation

Total RNA was extracted from bacterial cells by TRIzol reagent (Invitrogen, USA) according to the manufacturer's instructions with modifications. Briefly, bacterial cells were pelleted and homogenized in 1 ml of TRIzol reagent. The mixture was incubated at 65ºC for 30 min to achieve complete dissociation of nucleoprotein complexes. 0.2 ml of chloroform was added and the tubes were vigorously shaken with vortex for 15 sec and incubated at room temperature for 3 min. Samples were centrifuged at 13,000 x g for 15 min at 4°C. The aqueous upper phase was then transferred to a fresh tube and the RNA precipitated by mixing with isopropanol and 0.8M sodium citrate/1.2M NaCl, 0.5 volumes of the aqueous phases each. The sample was incubated at room temperature for 10 min and centrifuged at 14,000 x g for 20 minutes at 4°C. The supernatant was discarded and the RNA washed twice with 1 ml 75% ethanol and centrifuged at 14,000 x g for 5 min at 4°C. The RNA was then air-dried and dissolved in RNase-free water.

4.8.2 DNase reaction

This was essential to remove DNA contaminants from the RNA preparation.

Procedure

Reaction components	RNA (up to 1µg)	x µl
	DNase incubation buffer	1 µl
	DNase	1 µl
	RNase free water	ad to 10 µl
Incubation	15 min at RT	
Reaction stop	25 mM EDTA (pH 8)	1 µl
	10 min heat treatment at	

65°C

4.8.3 Reverse Transcription

Reverse transcription is an enzyme-catalyzed synthesis of cDNA from an RNA matrix in the presence of a gene specific primer and dNTPs. The SuperscriptTM II RNase H^- Reverse Transcriptase (Invitrogen, USA) was used in all reverse transcription analyses as described below. As a negative control reaction, RNA sample without reverse transcriptase was also always included to exclude the possibility of false positive reactions due to DNA contamination.

4.9 Bacterial transformation

Bacterial cells were made electrocompetent using standard procedures, and then transformed with plasmid/cosmid DNA as described below.

4.9.1 Production of electrocompetent cells

A modified protocol from Hanahan (Hanahan, 1983) was employed. The procedure was carried out in the cold and under sterile conditions.

Procedure

50 ml LB medium was inoculated with an overnight culture of the bacterium (*E. coli* or *Yersinia*) and incubated with vigorous shaking at 27°C/37°C until an OD_{600} of 0.5 - 0.6 was achieved.

- The cells were chilled on ice for 10 - 15 min and transferred to 50-ml falcon tubes.
- Cells were then centrifuged at 4000 rpm for 25 min at 4°C.
- The supernatant was decanted and cells resuspended in 50 ml of sterile ice-cold water (sterile), mixed well and centrifuged under the same conditions as above.
- The above wash step was repeated, following which cells were washed with 50 ml ice-cold 10% glycerol (centrifuged in the cold at 4000 rpm for 25 min).
- The glycerol solution was decanted and the cell volume estimated and cells resuspended in an equal volume of ice-cold glycerol.
- Cells were then aliquoted in 50 µl volumes and stored at -80°C until required.

Transformation

Electroporation with high voltage was achieved with the Gene Pulser II from Bio-Rad. The principle relies on the fact that short electrical impulses directed at bacterial cells generate

pores in the cell membrane that facilitates entry of foreign or exogenous DNA into the cell (Dower *et al*., 1988). The settings employed were 25 µF capacitance at 2.5 kV and 200 ohms After electroporation transformed cells were mixed with 1 ml LB medium and incubated at 27°C/37°C with shaking for 50 min. Bacterial cells were then plated out in 100 - 200 µl aliquots on LB-agar plates containing the required antibiotics for selection of recombinants.

4.10 Southern Blot hybridization (Southern, 1975)

This method was originally described by Southern and the principle involves transfer of DNA from a gel to a membrane, fixation of transferred DNA with UV light and subsequent hybridization of the membrane with an appropriately labeled DNA probe. The objective is to locate DNA samples that share some level of homology with the labeled probe. DNA probes were typically labeled with digoxigenin according to manufacturers' instructions (Roche, Mannheim).

4.10.1 Preparation of DNA probe

4.10.1.1 Digoxigenin-labeling of DNA through PCR

Procedure

Reaction components	Template DNA	2.5 µl
	PCR Buffer	5 µl
	dNTPs (2 mM dATP/dCTP/dGTP and 1.3 mM dTTP)	5 µl
	DIG dUTP	3.5 µl
	Primer 1 (100 pmol/µl)	1 µl
	Primer 2 (100 pmol/µl)	1 µl
	Taq-polymerase	0.2 µl
	H_2O_{dest}	31.7 µl
PCR amplification	35 PCR cycles of denaturation, annealing and elongation	

4.10.1.2 Random-primed method of DNA labeling

This method relies on the random labeling of a DNA sample with DIG-11-dUTP (Roche, Mannheim), catalyzed by the Klenow polymerase enzyme.

Procedure

Starting component	DNA cooked for 10 min and rapid-chilled on ice	10 ng – 1 µg
Other reaction components	Hexanucleotide mix (10x)	2 µl
	dNTP labeling mix (10x)	2 µl
	Klenow enzyme (2 U/µl)	1 µl
	H_2O_{dest}	20 µl
Incubation	37°C, overnight	
Further reaction components	0.2 M EDTA	2 µl
	4M LiCl	2 µl
	ice-cold C_2H_5OH 96%	50 µl
Incubation	-70°C, 30 min	

The labeled DNA was then pelleted by centrifugation at 15,000 rpm at 4°C for 15 min, the DNA pellet washed with ice-cold C_2H_5OH and in 50 µl H_2O_{dest} suspended.

4.10.2 Southern (Vacuum) Blot

Basically the Southern blot was carried out by transferring DNA from an agarose gel onto a nylon membrane using a vacuumblot (LKB 2016 Vacu Gene^R, Pharmacia-LKB, Uppsala, Sweden). The blotting involved treating the gel consecutively with a depurination solution, a denaturating solution, a neutralizing solution and finally 20 x SSC at a minimum pressure of 40 mbar for depurination and 50 mbar for the following steps. A successful DNA-transfer step was then followed by DNA fixation on a nylon membrane by crosslinking with UV light (0.12 J/cm^2).

Depurination solution	0.25 M HCl
Denturating solution	0.5 M NaOH; 1.5 M NaCl
Neutralizing solution	0.5 M Tris-HCl, pH 7.5; 1.5 M NaCl
20 x SSC (pH 7.0)	0.3 M Na_3 citrate; 3 M NaCl

4.10.3 Hybridization and detection

Hybridization solution	SSC	5x
	SDS	0.1%
	Blocking reagent (dry milk, Roth)	1%
Wash-buffer 1	SSC	2x

	SDS	0.1%
Wash-buffer 2	SSC	0.1x
	SDS	0.1%
Blocking solution	Blocking reagent in buffer 1	1%
Buffer 1 (pH 7.5)	Tris-HCl	0.1 M
	NaCl	0.15 M
Substrate-buffer (pH 9.6)	NaCl	0.1 M
	Tris-HCl	0.1 M
	$MgCl_2$	0.05 M
Substrate-detection solution	NBT (1 mg/ml)	10 ml
	BCIP (5 mg/ml)	1 ml
	Substrate buffer	90 ml

Hybridization: The membrane blot was first prehybridized for a minimum of 3 hours at 68°C, followed by hybridization with the appropriate DIG-probe at 68°C overnight. Stringency washes then followed comprising two consecutive washes of the blot at RT for 15 min with wash buffer 1, followed by another round of two stringency washes at 68°C for 30 min each with wash buffer 2 (see below for buffer compositions).

Blocking and incubation of the blot with Anti-Digoxigenin antibody: The membrane blot was then incubated with blocking buffer for 1 hr at RT and then washed briefly with buffer 1. This step prevented unspecific binding of the DNA probe to non-homologous DNA regions. The membrane was then washed briefly with buffer 1, followed by incubation for 45 min of the membrane blot in a solution containing the Anti-Digoxigenin antibody conjugated with alkaline phosphatase (the Anti-Dig antibody was diluted 1:5000 in buffer 1). The membrane was then washed twice with buffer 1.

Detection: Detection followed, by incubation in a freshly mixed substrate-detection solution at 37°C in the dark. On visible detection of the bands, the reaction was stopped by the addition of water.

4.11 Cosmid gene bank of *Y. enterocolitica* Y-108C

The cosmid gene library was prepared with the Supercos 1 cosmid vector (Stratagene). This is a vector specially engineered to contain bacteriophage promoter sequences (*cos* sites) flanking a unique cloning site, thus allowing the in vitro packaging of DNA into phage heads.

4.11.1 Preparation of cosmid vector DNA

- 25 µg of the cosmid vector DNA (7.6 kb) was digested with 9 U/µg of *Xba*I in a total volume of 200 µl according to standard digestion procedures (see 4.6.1) for 1 hr at 37°C. Digested vector was purified once with phenol-chloroform-isoamyl alcohol (25:24:1). The *Xba*I digested cosmid DNA was resuspended in distilled water at a concentration of 1 µg/µl and subjected to dephosphorylation according to previously described protocols (see 4.6.2).
- The dephosphorylated DNA was phenol-purified and digested with 5 U/µg of *Bam*HI endonuclease in a total volume of 200 µl at standard buffer conditions. Complete digestion was verified by loading DNA sample on 0.8% agarose gel. Two cosmid bands were observed running at 1.1 and 6.5 kb respectively.
- The X*ba* I/*Bam*H I digested cosmid DNA was purified with phenol-chloroform, resuspended in deionized water at a concentration of 1 µg/µl and stored at -20°C.

4.11.2 Preparation of genomic DNA

Procedure

- Isolation of chromosomal DNA was according to the CTAB/NaCl procedure previously described (see 4.1).
- The chromosomal DNA was partially digested with *Sau*3AI in order to clone into the *Bam*HI site of the Supercos 1 cosmid vector. Preliminary digests prior to the main one were carried out to identify the optimal incubation time to achieve an insert size range of 30 - 42 kb, essential for cloning into the Supercos1 vector.
- Preliminary digest: 10 µg of genomic DNA in a total reaction volume of 100 µl was digested at 37°C with 0.5 U of *Sau*3A I. 15 µl aliquots of the digest was then removed at various time intervals: 0 -, 5 -, 10 -, 15 -, 20 -, 30 -, and 45 minutes time points, and checked by gel electrophoresis on a 0.6 - 0.7% (w/v) agarose gel. The time interval that yielded a large proportion of the digested fragments running within 35 - 50 kb range was chosen as the desired digestion time.
- Main digest**:** After the optimal time interval was determined (15 min), a partial *Sau*3AI digest of 100 µg of genomic DNA in a 1 ml total reaction volume was carried out. The reaction was scaled up to best mimic the test partial digest including enzyme concentration, temperature and reaction time. The reaction was stopped with 15 µl of 0.5

M EDTA (pH 8). 10 µl aliquot of the reaction mix was loaded on a 0.7% gel to ensure appropriate size distribution. The DNA was resuspended in 50 µl of TE buffer.

- Dephosphorylation of the partially digested genomic DNA with SAP (see 4.6.2) after which it was phenol-chloroform purified. For dephosphorylation, the following were added to the 50 µl of DNA from above:

10x SAP buffer	10 µl
Distilled water	38 µl
SAP	2 µl (1 unit/µl)

The reaction volume was made up to 100 µl, and incubated at 37°C for 1 hr.

- 3 µl of 0.5 M EDTA was added to stop the reaction followed by heat inactivation of the enzyme at 65°C for 20 min.
- The DNA was extracted once with phenol-chloroform saturated with 50 mM Tris-HCl (pH 8.0) and once with chloroform. The aqueous phase was then adjusted to 0.3 M sodium acetate (pH 5.5) and ethanol-precipitated by adding 2.5 volumes of 100% (v / v) ethanol.
- The DNA was resuspended to a 1µg/µl concentration in TE buffer and a sample run on 0.7% agarose gel.

4.11.3 Ligation and packaging of DNA

- The ligation reaction was set up by adding the following reaction components to a microcentrifuge tube:

2.5 µg of partially digested SAP genomic DNA

1.0 µl of Supercos 1 DNA (*Xba* I-SAP and digested with *Bam*H I, 1 µg/µl)

2.0 µl of 10x ligase buffer

Water to a final volume of 20 µl

- A negative control ligation was set up by adding the following components to a microcentrifuge tube:

1.0 µl of Supercos 1 DNA (*Xba* I-SAP and digested with *Bam*H I, 1 µg/µl)

2.0 µl of 10x ligase buffer

Water to a final volume of 20 µl

- 1 µl aliquot was removed from each reaction and stored at 4°C for later gel analysis.
- 1 µl of T4 DNA ligase was added to the remaining 19 µl of the reaction and incubated at 4°C overnight.

- 1 µl aliquot was removed from each reaction and loaded on a 0.7% agarose gel for analysis.

Packaging

The packaging reaction was carried out with the Gigapack II gold packaging extract (Stratagene) according to manufacturers' instructions. *E. coli* XL 1- Blue MR was the bacterial host strain utilized for titration of the cosmid packaging reaction.

4.12 Suppressive subtractive hybridization (Diatchenko *et al*., 1996)

Genomic differences between *Y. enterocolitica* 4 (Y-108C, weakly pathogenic strain) and *Y. enterocolitica* 1B WA-314 (high-pathogenic strain) as well as *Y. pestis* Yokohama and *Y. pestis* G8786 were mapped out by employing the PCR-Select Bacterial Genome Subtraction System from Clontech (BD Biosciences Clontech, USA). An overview of the PCR-Select method is shown in Figure 1. Briefly, genomic DNA (1.5 - 2 µg) was first isolated from the two strains of bacteria to be compared (driver strain and tester strain). The chromosomal DNA was then digested with *Rsa*I, a four base cutter. This step generated DNA fragments ranging in size from 0.1 to 3 kb. The tester DNA was then subdivided into two portions, each of which was ligated with a different oligonucleotide adaptor (adaptor 1 and 2R at 10 µM concentration each). After adaptor ligation, the DNA was purified by phenol/chloroform extraction.

4.12.1 Hybridization

4x hybridization buffer	200 mM HEPES-HCl (pH 8.0)
	2 M NaCl
	0.2 mM EDTA (pH 8.0)

- For the first round of subtractive hybridization, approximately 0.6 µg of *Rsa*I-digested driver DNA was mixed in two separate reactions with 0.012 µg each of i) adaptor 1-ligated tester DNA and ii) adaptor 2R-ligated tester.
- The reaction volume was made up to 4 µl using a 4 x hybridization buffer. The samples were covered with a drop of mineral oil, centrifuged briefly and incubated in a thermal cycler (Perkin Elmer) at 98°C for 1.5 min (initial denaturation).

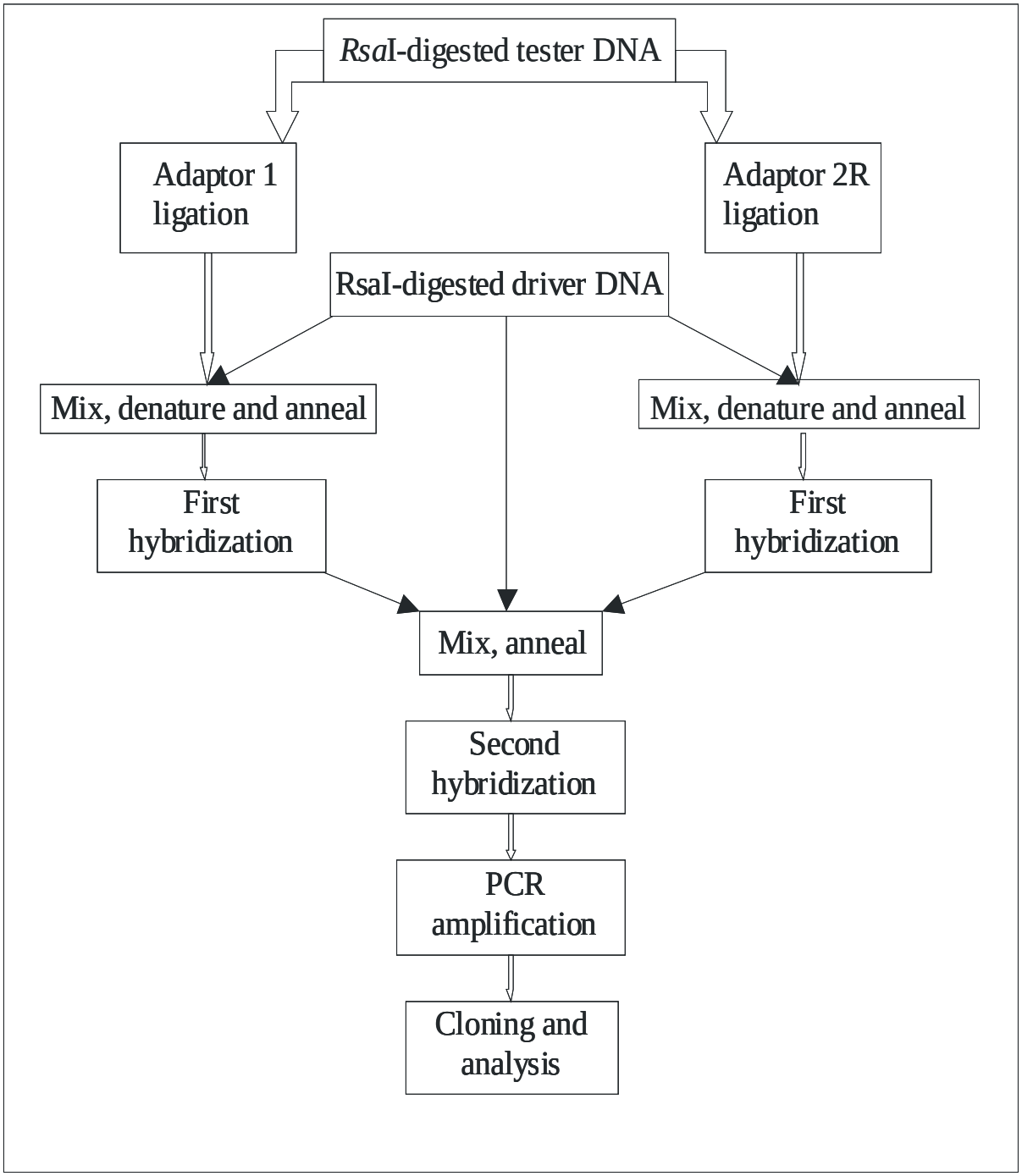


Fig. 1: Overview of the CLONTECH PCR-Select procedure.

Genomic DNA (1.5 - 2 µg) was first isolated from the two strains of bacteria to be compared (driver strain and tester strain). The chromosomal DNA was then digested with *Rsa*I, a four base cutter. This step generated DNA fragments ranging in size from 0.1 to 3 kb. The tester DNA was then subdivided into two portions, each of which was ligated with a different oligonucleotide adaptor (adaptor 1 and 2R at 10 µM concentration each). After adaptor ligation, the DNA was purified by phenol/chloroform extraction.

- Incubation followed at 63°C for 1.5 hr. The two samples from this first hybridization were then mixed together and 0.3 µg freshly denatured driver DNA was added to further enrich for tester-specific sequences.
- This DNA mixture was then subjected to a second hybridization reaction, with overnight incubation in the thermal cycler at 63°C.

4.12.2 PCR amplification

Prior to thermal cycling, a preincubation step was carried out at 72°C to fill in the missing strands of the adaptors and thus create binding sites for the PCR primer. This would provide binding sites for PCR Primer 1 whose nucleotide sequence corresponds to the first 22 nucleotides of both adaptors 1 and 2R. PCR amplification to selectively amplify for tester-specific DNA was done in two stages. In the first amplification, only dsDNAs with different adaptor sequences on both ends are sequentially amplified. In the second round of amplification, nested PCR, employing nested primers 1 and 2, was used to further reduce the background and enrich the DNA pool for tester specific sequences. The nucleotide sequences of the two nested primers 1 and 2 corresponded to the last 22 and 20 nucleotides respectively of Adaptors 1 and 2R utilized in the DNA ligation reaction.

Cycling parameters

Primary PCR	Denaturation	94°C	30 sec
	*Denaturation	94°C	30 sec
	*Annealing	66°C	30 sec
	*Elongation	72°C	1.5 min
	* 25 cycles		
Secondary nested PCR	Denaturation	94°C	30 sec
	*Denaturation	94°C	30 sec
	*Annealing	68°C	30 sec
	*Elongation	72°C	1.5 min
	* 10-12 cycles		

Analysis of the subtracted fragments on a 1.2% agarose gel revealed a smear with distinct bands running from about 300 bp to 1.5 kb. The subtracted fragments were subsequently cloned into pMOS*Blue* vector (Amersham Biosciences, USA), followed by transformation into the highly efficient JM109 competent cells (see section 4.9 for transformation

procedure). The pMOS*Blue* vector allows for blue-white screening with recombinant clones appearing white when plated on X-gal and IPTG indicator plates.

4.12.3 Preparation of X-gal/IPTG LB-agar plates for blue-white screening of recombinants

- For one plate 35 µl of 50 mg/ml X-gal and 20 µl of 100 mM IPTG were added to 30 ml LB-agar with an appropriate antibiotic.
- The medium was dropped on plates.
- The plates were left to soak for at least 30 min prior to plating.
- 10 - 50 µl of each transformant was then spread on the LB agar X-gal/IPTG plates.

Inverted plates were incubated overnight at 37°C.

5. Protein biochemical studies

5.1 Sodium-dodecyl-sulphate Polyacrylamide Gel Electrophoresis

Principle

In SDS polyacrylamide gel electrophoresis, proteins are separated as they migrate through a gel on the basis of their molecular weights. SDS is an anionic detergent that denatures proteins. The SDS also disrupts hydrogen bonds, blocks hydrophobic interactions, and substantially unfolds the protein molecules by eliminating the tertiary and secondary structures. Two types of buffer systems are used in protein gel electrophoresis: continuous and discontinuous. In the discontinuous system employed in this work, a non-restrictive large-pore gel called a stacking gel is layered on top of a separating (resolving gel). The buffer composition for the two gel layers differs which in turn differs from the composition of the electrophoresis buffer. At the onset of an electrophoretic separation, the proteins migrate first through the stacking gel and then into the separating gel, where separation takes place. With the aid of a protein marker applied alongside the protein samples of interest, the MW of the proteins applied on the gel can be estimated. The following is the pipetting scheme applied for the preparation of two 11% acrylamide SDS-gels:

Separating gel	**Stacking gel**
2.15 ml H_2O	3.19 ml H_2O
3.75 ml 1 M Tris/HCl (pH 8.8)	0.83 ml 0.75 M Tris/HCl (pH 6.8)
3.7 ml Protogel	0.7 ml Protogel

0.2 ml SDS (10%)	0.1 ml SDS (10%)
40 µl APS (10%)	40 µl APS (10%)
200 µl TEMED (10%)	200 µl TEMED (10%)

10x Electrophoresis buffer

Tris	30.2 g
Glycine	142.6 g
H2O	ad 1 liter

4x SDS-loading buffer (pH 6.8)

Tris	0.4 g
SDS	1.2 g
Glycerol	7.5 ml
ß-Mercaptoethanol	2.5 ml
Bromophenol blue (2% solution)	0.5 ml
H_2O	ad 50 ml

Coomasie dye solution

0.125% Coomasie Brilliant blue (Serva)	250 mg

Destaining solution

Methanol	250 ml
Glacial acetic acid	350 ml
H_2O	ad 5 liter

The electrophoresis system from Bio-Rad was employed in this work and the assembly of glass plates and spacers for the production of the gels was according to manufacturer's instructions. For the SDS-gel run, the probes to be analyzed were mixed with the SDS-loading gel buffer and cooked briefly at 95°C for 5 - 10 min and then applied on the gels. Electrophoresis proceeded at an applied voltage of 200 V (or at 20 mA) for 1 - 2 hr.

5.2 Western Blot (Towbin *et al*., 1979)

Principle

Western blot relies on the principle of the specificity of the interaction between a protein and its cognate antibody which is visualizable by means of chemiluminiscence or autoradiography.

Reagents

1x Western-blot buffer	0.025 M Tris
	0.19 M Glycine
Wash buffer	0.5% Tween 20 (v/v) in 1x PBS
Blocking solution	5% milk (w/v) in wash buffer

Procedure

- Following separation of proteins in a conventional SDS-PAGE, the proteins were electrophoretically transferred to a nitrocellulose membrane (in 1x Western blot buffer) at a constant voltage of 150 - 200 V (or 350 mA) for 1 hr.
- The membrane was then incubated in blocking solution overnight in the cold.
- The membrane was then incubated for 1.5 hr with the first antibody directed against the protein of interest, following which the blot was washed three times.
- The secondary antibody, conjugated with either alkaline phosphatase (AP) was then added (usually at a 1:5000 dilution), incubation for 1.5 hr followed, after which the membrane blot was washed three times.
- The detection of protein-antibody interaction or binding was by means of the BCIP/NBT-Blue liquid substrate system (Sigma) for alkaline phosphatase-coupled secondary antibody.

The BCIP/NBT system is a colorimetric antigen detection method. NBT and BCIP are two colorless substrates which form a redox system. BCIP is oxidized by alkaline phosphatase to indigo by release of a phosphate group. In parallel, NBT is reduced to diformazan. The reaction products form a water insoluble brownish precipitate on nylon membranes.

5.3 Cultivation and induction of bacteria

E. coli BL21 carrying the pET100/D-TOPO/RtxA fusion vector was cultivated at 37°C overnight in LB medium fortified with ampicillin. The culture was diluted 1:100 in LB-medium (containing ampicillin) and incubated at 37°C with shaking till an OD_{600} of 0.6 - 0.8 was achieved. The cells were then induced with 0.3 mM IPTG, incubated further at 16°C for 32 hr. The cells were then pelleted by centrifugation at 6000 rpm for 20 min at 4°C, and the pellet resuspended in PBS (containing 1 mM PMSF, 1 mM DTT for protein stabilization). For

release of the soluble protein fractions from the cell, the bacterial suspension from above was subjected to French Press with the French Pressure Cell at 1000 psi (repeated four times).

5.4 Purification of the 6xHis fusion protein

The soluble fractions with the 6xHis fusion protein from bacterial lysates were rapidly purified with Ni-NTA Purification System (Invitrogen, USA). The principle is based on the strong affinity of the polyhistidine (6xHis) peptide to nickel-charged agarose resins which it binds specifically; allowing other proteins to flow though the column packed with the agarose beads. Through several wash steps, the unspecific bound proteins are washed through the column and the 6xHis-tagged protein eluted under mild conditions with an elution buffer.

Procedure

Cell debris was dissolved in loading buffer (6 M Urea, 40 mM imidasole, 20 mM Tris - HCl, pH 8.0) and applied on a 3-ml Ni-NTA column (Bio-Rad) equilibrated with the same buffer. After loading, the column was washed with 10 bed volumes of loading buffer. 6xHis-fusion protein was produced by elution from column under denaturating conditions (6M Urea, 200 mM imidasole 20 mM Tris - HCl, pH 8.0). Eluted protein was collected in 0.4 ml portion and frozen at -30°C. Protein sample was analyzed by SDS-PAGE and Western-blot with AP Ni – NTA conjugate (Qiagen).

5.5 Preparative SDS-PAGE and protein recovery

5.5.1 Preparative SDS-PAGE

The fusion protein was mixed with SDS-loading gel buffer (Laemmli) and loaded on a preparative SDS-Polacrylamide gel. The run conditions were as described previously. Following the SDS-PAGE run, the gel was stained with Coomasie dye solution and the required protein band was excised from the gel with a clean scalpel.

5.5.2 Protein recovery

Gel slice containing the protein was homogenized and macerated with 300 µl of PBS buffer and crushed using a mortar and pestle. Homogenous suspension was used for a rabbit immunization as described below.

5.6 Rabbit immunization

The rabbit was injected three times subcutaneously, with a time period of three weeks between each immunization. Typically, 400 - 500 µg of protein was used per immunization. For the first immunization, the protein was mixed with an equal proportion of complete Freund's adjuvant. Further immunizations utilized Freund's incomplete adjuvant also in a 1:1 ratio with the protein. Ten days after every immunization, blood samples were collected from the animal. The blood was allowed to coagulate at RT for 3 - 5 hours and centrifuged for 30 min at 6000 rpm. The serum (supernatant) was then collected and stored at -20°C.

5.7 Immunoprecipitation (Cochet *et al*., 1998)

Immunoprecipitation was carried out using Dynabeads^{R} Protein G (DYNAL BIOTECH ASA, Oslo, Norway). Dynabeads^{R} Protein G are uniform, magnetizable polystyrene beads 2.8 µm in diameter and coated with recombinant protein G covalently coupled to the surface.

Procedure

- 50 µl of the Dynabeads^{R} Protein G were added to the serum which contains anti RtxA79 IgG.
- The suspension was incubated with gentle mixing for 40 min at RT.
- The Dynabeads^{R} Protein G/IgG complex was washed three times in the PBS/0.1% Tween-20 buffer by using the magnetic PickPen^{R} (BioNobile, Finland).
- The Dynabeads^{R} Protein G/IgG complex was added to the bacterial lysate containing RtxA and incubated with tilting and rotation at 4°C for 3 hr.
- The protein-beads complex was collected with PickPen^{R} and washed three times by using 1 ml of PBS/0.1% Tween-20 buffer.
- RtxA was eluted from the magnetic beads with the 40 µl of Laemmli sample buffer (SDS-PAGE sample buffer) at 95°C for 10 min and then applied on the gel for analysis.

6. Bioinformatics

Bioinformatic tools were powerfully utilized for sequence analysis, alignments and similarity searches. The two primary databanks that were extensively utilized were Genbank and EMBL (European Molecular Biology Laboratory).

Primary databanks

1. The Genbank in the USA is under the patronage of the National Center for Biotechnology (NCBI) and is an official Sequence data bank which contains more than 3 millions protein

and nucleotide sequences. All sequences are identified or tagged with a unique accession number. A Genbank sequence is usually divided into two parts:

- the Annotation which contains a precise and detailed information about the sequence and
- the Sequence itself. The ENTREZ search machine is coupled with the Genbank and allows a specific search based on an accession number, organism, gene, protein or author.

2. The EMBL nucleotide sequence database is the European equivalent of the Genbank and utilizes the SRS (sequence retrieval system), a search machine similar to the ENTREZ for specialized searches of the database and many other databanks over the web interface.

BLAST

In addition to the text-based SRS and ENTREZ search engines described above, the BLAST search was also extensively utilized. The BLAST (basic local alignment search tool) search enables comparison of a particular sequence of interest with available databanks, leading to identification of similar sequences or relationships with previously described gene families. The following BLAST programs were employed in this work:

- BLASTN: compares a nucleic acid query sequence with nucleic acid databanks directly
- BLASTX: compares a translated nucleotide sequence with protein sequence databanks
- TBLASTX: compares a translated nucleotide sequence with a database of translated nucleotide sequences
- BLASTP: compares a protein query with a protein database.

The BLAST program provided by NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) and BLAST2 (http://www.ch.embnet.org/software/BottomBLAST.html?) maintained by the Swiss Institute of Bioinformatics were extensively used for sequence analysis.

FASTA: A very common format for sequence data is derived from conventions of FASTA, a program for FAST Alignment by W. R. Pearson. Many of the programs used in this work employed the FASTA format for reading sequences or for reporting results.

Sequence alignment: This is the assignment of residue-residue correspondences. Examples included:

- a Global match: all of one sequence was aligned with all of another
- a Local match: a region in one sequence was matched with a region of another
- a Multiple alignment: a mutual alignment of many sequences.

7. Nucleotide sequence accession numbers

The sequences of the *Y. enterocolitica* O:3 IS*1331A*, IS*1331B* insertion elements and the pG8786 plasmid from *Y. pestis* G8786 have been submitted to the EMBL database under the accession numbers AJ849657, AJ849658 and AJ698720, respectively.

C. RESULTS

1. Uncovering genomic differences in human pathogenic *Yersinia enterocolitica*

1.1 Construction of libraries of subtracted fragments and their analysis for tester-specific sequences

To uncover genomic differences between mouse virulent and mouse non-virulent human pathogenic yersiniae, we subtracted genomic DNA of the mouse virulent WA-C, BG 1B, ST O:8 strain with genomic DNA of mouse non-virulent Y-108C, BG 4, ST O:3 strain using WA- C strain first as a tester and than as a driver. Both, O:8- and O:3-specific sequences were recognized by hybridization. Only DNA fragments with possible impact on virulence and those, which can serve as epidemiological markers are depicted in Tables 7 and 8. The percentages of clones in the subtracted libraries that correspond to tester-specific DNA were 53% and 49% for O:8 and O:3, respectively (checked by Southern Blot hybridization, data not shown). In total, 428 O:8-specific and 83 O:3-specific sequences were identified. Not surprisingly, 60% of O:8- and 44% of O:3-specific sequences showed high similarity (70-95%) to the sequences of the completed *Yersinia pestis* CO92 genome (Tables 7 and 8).

Table 7. Selected *Y. enterocolitica* O:3-specific SSH fragments and their characteristics

No.	Homology to predicted gene and encoded protein	Similarity, positives (% amino acid)	GenBank accession number
01_05	*wbbX*, lipopolysaccharide O-antigen biosynthesis protein, *Yersinia enterocolitica* O:3	162/162 (100%)	CAA79354
	wbbW, lipopolysaccharide O-antigen biosynthesis protein, *Yersinia enterocolitica* O:3	71/71 (99%)	CAA79353
01_22	YPO0732, putative flagellar hook-associated protein, *Yersinia pestis*	130/215 (59%)	CAC89583
01_27	YPO0741, putative flagellar protein, *Yersinia pestis*	56/72 (77%)	CAC89592
01_33	*xnp2*, nematocidal toxin, *Xenorhabdus bovienii*	139/275 (49%)	CAC19493
01_43	YPO2274, putative phage protein, *Yersinia pestis*	198/204 (96%)	CAC91078
01_52	*rtxA*, cytotoxin, *Vibrio cholerae*	150/192 (77%)	AAD21057
01_65	YPO0466, hypothetical protein, *Yersinia pestis*	144/188 (75%)	CAC89322
01_92	YPO0272, putative type III secretion apparatus protein, *Yersinia pestis*	61/87 (69%)	CAC89135
	wbbT, lipopolysaccharide O-antigen biosynthesis protein,	141/143 (98%)	CAA79348

	Yersinia enterocolitica O:3		
02_11	*yapG*, putative autotransporter protein, *Yersinia pestis*	55/67 (81%)	CAC14226
02_31	YPO0261, putative type III secretion apparatus, *Yersinia pestis*	25/47 (52%)	CAC89123
	YPO0262, putative type III secretion apparatus, *Yersinia pestis*	58/83 (69%)	CAC89124
	YPO0263, type III secretion system apparatus lipoprotein, *Yersinia pestis*	45/50(90 %)	CAC89125
02_33	YPO2279, putative phage-related membrane protein, *Yersinia pestis*	193/199 (96%)	CAC91084
02_64	*sseD*, secretion system effector, *Salmonella typhimurium*	45/81 (55%)	AAL20325
02_72	*yapA*, putative autotransporter protein, *Yersinia pestis*	79/98 (80%)	CAC14220
02_85	YPO2884, putative exported protein, *Yersinia pestis*	112/197 (56%)	CAC92135
02_92	YPO2273, phage hypothetical protein, *Yersinia pestis*	29/43 (66%)	CAC91077
	sscA, secretion system chaperone, *Salmonella typhimurium*	63/105 (59%)	AAL20323
02_93	YPO0720, putative flagellar regulatory protein, *Yersinia pestis*	37/50 (74%)	CAC89571
	YPO0719, hypothetical protein, *Yersinia pestis*	78/107 (72%)	CAC89570
03_10	*wbcM*, putative glycosyltransferase, *Yersinia enterocolitica* O:3	50/68 (73%)	CAA87701
03_28	VCA0253, antibiotic acetyltransferase, *Vibrio cholerae*	56/95 (58%)	AAF96164
03_33	ORF, putative multidrug-resistance protein, *Aeromonas hydrophila*	63/147 (42%)	AAF63418

Table 8. Selected *Y. enterocolitica* O:8-specific SSH fragments and their characteristics

No.	Homology to predicted gene and encoded protein	Similarity, positives (% amino acid)	GenBank accession number
04_62	*EcoR*IIM, site-specific DNA-methyltransferase (cytosine-specific), *Escherichia coli*	198/222 (88%)	P05101
04_37	trp*1400B*, IS*1400* transposase B, *Yersinia enterocolitica* O:8	91/92 (98%)	CAB46580
04_50	trp*1329B*, IS*1329* transposase, *Yersinia enterocolitica* O:8	148/148 (100%)	CAB46576
04_93	YPO2884, putative exported protein, *Yersinia pestis*	29/44 (65%)	CAC92135
05_02	ECs5005, hypothetical protein, *E. coli*	79/92 (85%)	NP_313032
05_04	YPO2999, putative HlyD secretion protein, *Yersinia pestis*	140/166 (84%)	NP_406493
05_19	*yen*I, methyltransferase-endonuclease, *Yersinia enterocolitica* O:8	39/39 (100%)	CAC95150
05_51	YPO0466, hypothetical protein, *Yersinia pestis*	89/179 (49%)	CAC89322
05_76	YPO3268, multidrug resistance protein B, *Yersinia pestis*	68/93 (72%)	CAC92502
05_84	VCA0811, putative chitinase, *Vibrio cholerae*	83/165 (49%)	AAF96709

06_31	YPO3133, multidrug efflux protein, *Yersinia pestis*	135/136 (98%)	CAC92368
06_37	*cya*, adenylate cyclase, *Yersinia enterocolitica* O:8	31/32 (96%)	AAL51150
06_42	tpn*1330*, transposase, *Yersinia enterocolitica* O:8	197/200 (98%)	CAC83041
06_53	*hmsR*, putative hemin binding protein, *Yersinia pestis*	41/42 (97%)	T47005
	hmsS, hemin storage system, HmsS protein, *Yersinia pestis*	50/56 (88%)	CAC90769
06_87	*shuA*, outer membrane heme receptor, *Shigella dysenteriae*	65/145 (44%)	AAC27809
07_14	*fes*, enterochelin esterase, *Yersinia enterocolitica* O:8	86/89 (96%)	AAB02191
07_92	*pfeA*, ferric enterobactin receptor precursor PfeA, *Pseudomonas aeruginosa*	69/94 (72%)	AAG06076
07_96	*fcl*, putative GDP-L-fucose synthetase, *Yersinia enterocolitica* O:8	157/165 (95%)	AAC60774
08_04	YPO0744, putative flagellar biogenesis protein, *Yersinia pestis*	64/91 (69%)	CAC89595
08_08	*manC*, GDP-mannose pyrophosphorylase, *Yersinia enterocolitica* O:8	233/248 (93%)	AAC60775
08_32	*wbcG*, putative glycosyltransferase, *Yersinia enterocolitica* O:8	169/170(99%)	AAC60770
08_62	*pic*, protease with mucinase and hemagglutinin activities, *Shigella flexneri*	67/134 (49%)	AAK00464
12_18	orf, ferrichrome-iron receptor, *Zymomonas mobilis*	26/52 (49%)	AAG42412
12_34	*cnf*, cytotoxic necrotizing factor, *Yersinia pseudotuberculosis*	64/140(45%)	AAG45433
13_40	YPO1496, putative heme-binding protein, *Yersinia pestis*	192/229 (83%)	CAC90319
13_95	YPO3922, hemophore HasA, *Yersinia pestis*	109/139 (77%)	CAC93387

1.1.1 Subtracted fragments with similarity to genes of surface structures and metabolic pathways

Genes involved in synthesis of O:3-specific (*wbbX*, *wbbW*, and *wbbT*) (Zhang *et al*., 1993) and O:8-specific (*fcl*, *manC*, and *wbcG*) (Zhang *et al*., 1997) O-antigens, as well as host-specific restriction modification systems, *Eco*RII-like (Som *et al*., 1987) in *Y. enterocolitica* O:3 and *Yen*I, *Pst*I-like in *Y. enterocolitica* O:8 were uncovered as species-specific (Kinder *et al*., 1993). Sequences with similarities to genes of the *Y. pestis* hemin storage locus (*hmsR*, *hmsS*) (Buchrieser *et al*., 1999; Parkhill *et al*., 2001) were found only in *Y. enterocolitica* serotype O:8, although *Y. enterocolitica* does normally not express the pigmentation phenotype of *Y. pestis*. On the other hand, the sequences of one of the two flagellar operons, found in *Y. pestis* CO92 (namely YPO0719, YPO0720, YPO0732, YPO0741, and YPO0743) (Parkhill *et al*., 2001), were absent from O:8 but present in strain O:3. Besides the

yersiniabactin iron uptake and hemin storage systems, DNA of strain O:8 contained sequences with similarities to systems that are involved in iron and heme (hemin) acquisition, such as Fes enterochelinesterase (Schubert *et al*., 1999), PfeA ferric enterobactin receptor from *Pseudomonas aeruginosa* (Stover *et al*., 2000), or YPO1496 putative heme binding protein and hemophore-dependent heme acquisition system Has from *Y. pestis* (Rossi *et al*., 2001), ferrichrome iron receptor from *Zymomonas mobilis*, and ShuA outer membrane heme receptor of *Shigella dysenteriae* (Mills and Payne, 1995; Koebnik *et al*., 1993).

1.1.2 Subtracted fragments with similarity to virulence factors

Certain candidates for virulence factors were revealed in both strains. Sequences with similarity to RtxA-like cytotoxin of *Vibrio cholerae* (Lin *et al*., 1999), type III secretion system of *Salmonella typhimurium* located on a SPI2 pathogenicity island (YscT, SseD, and SscA) (Nikolaus *et al*., 2001; Cornelis, 2002), Xnp2 nematocidal toxin from *Xenorhabdus*, or YapA and YapG putative autotransporter proteins from *Y. pestis* (Parkhill *et al*. , 2001), were specific for *Y. enterocolitica* O:3. Surprisingly, sequences with similarity to the proteins of the chromosomally encoded type III secretion system (TTSS) were more closely related to the putative proteins YPO0272, YPO0261, YPO0262 and YPO0263 of the *Y. pestis* CO92 than to the TTSS of *Y. enterocolitica* O:8 strain 8081 (Haller *et al*., 2000). On the other hand, sequences with similarity to putative HlyD secretion protein YPO2999 from *Y. pestis*, adenylate cyclase (Petersen and Young, 2002), Pic protease with mucinase and hemagglutinin activities from *Shigella flexneri* (Al Hasani *et al*., 2001), or CNF cytotoxic necrotizing factor found in *Yersinia pseudotuberculosis* (Lockman *et al*., 2002) turned out to be *Y. enterocolitica* O:8-specific sequences.

1.1.3 Subtracted fragments with similarity to drug resistance genes

Strains of both serotypes contain sequences that might be involved in drug resistance. VCA0253 streptogramin A acetyltransferase from *V. cholerae* (Seoane and Garcia Lobo, 2000) and putative multidrug resistance protein from *Aeromonas hydrophila* (Zhang *et al*., 2000) were revealed in *Y. enterocolitica* O:3 strain, while three proteins, YPO1267, a probable drug-resistant translocator protein, YPO3268, multidrug resistance protein B, and YPO3133, multidrug efflux protein from *Y. pestis* (Parkhill *et al*., 2001), were specific for *Y. enterocolitica* O:8.

1.1.4 Subtracted fragments with similarity to movable genetic elements

SSH readily mapped known *Y. enterocolitica* 1B-specific insertion sequences such as IS*1326*, IS*1327*, IS*1328*, IS*1329*, and IS*1440* (Rakin and Heesemann, 1995; Rakin *et al*., 2000), which might serve as epidemiological markers of *Y. enterocolitica* 1B group. At least some genes of the hypothetical prophage encoded by YPO2273 - 2279 of *Y. pestis* CO92 were present in *Y. enterocolitica* O:3 isolate, while a sequence with similarity to that encoding the Ecs5005 hypothetical protein (encoded by VT2-Sakaj prophage carrying the verotoxin 2 genes) (Makino *et al*., 1999) was restricted to *Y. enterocolitica* O:8 strain.

1.1.5 Subtracted fragments with similarity to genes of hypothetical proteins

The same *Y. pestis* sequences, YPO0466 and YPO2884, encoding hypothetical proteins (Parkhill *et al*., 2001) were presented in subtractive libraries of both strains. Still, YPO0466 sequence showed 77% homology with *Y. enterocolitica* O:3, but only 49% with *Y. enterocolitica* O:8 sequence. YPO2884 sequence demonstrated 56% similarity with that revealed in *Y. enterocolitica* O:3, but only a small fragment of YPO2884 (44 aa) had 65% homology with the sequence from *Y. enterocolitica* O:8 strain. Moreover, both YPO0446 and YPO2884 showed 77% homology with each other, emphasizing a common origin of both sequences.

For our further investigation we have chosen two subtractive fragments from *Y. enterocolitica* ST O:3 with similarity to a novel IS21-like element and putative RtxA-cytotoxin. Preliminary data have shown that IS21-like element might be unique for European weakly pathogenic *Y. enterocolitica* isolates. Putative RtxA-like cytotoxin might be the new virulence determinant of *Y. enterocolitica* ST O:3 strains.

2. A novel IS*21*-like element - IS*1331* was uncovered by subtractive hybridization

2.1 General description of IS*1331*

Suppression subtractive hybridization (SSH) was performed to distinguish the weakly pathogenic *Y. enterocolitica* BG 4 ST O:3 Y-108C strain from the highly pathogenic *Y. enterocolitica* BG 1B ST O:8 WA-314C strain and to identify genes specific to the weakly pathogenic *Y. enterocolitica* strain (Golubov *et al*., 2003). The variety of unique sequences was uncovered by SSH. Two fragments (03_20 and 01_46) of 83 Y-108C-specific sequences

have significant homology to genes of previously described transposases and NTP-binding proteins, respectively. A cosmid pSC1046 that hybridized with the original DIG-labelled 03_20 subtracted fragment was isolated from the *Y. enterocolitica* Y-108C cosmid gene library. Further sequencing of the cosmid by primer walking identified a putative insertion element, IS*1331*, which turned out to be unique to European weakly pathogenic *Y. enterocolitica* isolates.

Sequence comparison of all IS*1331* copies in *Y. enterocolitica* Y-108C ST O:3 (pYV-bearing strain), *Y. enterocolitica* gk 1142 O:2 and *Y. enterocolitica* JD E029 O:1 revealed two IS*1331* isoforms – IS*1331A* and IS*1331B*, which have differences in their nucleotide content and structure (Fig. 2, 3C).

In general, IS*1331* has a characteristic IS structure with two imperfect terminal inverted repeats (IRL and IRR) of 30 bp with eight mismatches (Fig. 2, 3AB). Comparison of the terminal inverted repeats of all IS*1331* copies has shown striking similarities among them (Fig. 3A, 3B). IS*1331* contains two long ORFs, the *istAB* genes that are organized in an operon (Fig. 2). These ORFs have been designated *istA* and *istB* based on the similarity of their predicted products to the putative transposase IstA (56% at the amino acid level, Fig. 4) and the ATP binding protein

IstB of IS*1326* (71% at the amino acid level, Fig. 5) (Brown *et al*., 1996). The GC content of the IS*1331* is 39.73%, which is not in agreement with the GC content of the *Y. enterocolitica* host (46% to 48% for the *Yersinia* chromosome) (Bercovier and Mollaret, 1984).

The main differences between two isoforms of IS*1331* are one 6 bp deletion in *istA* (Fig. 3C), and 14 single nucleotide replacements over IS*1331* sequence which changed the *istB* frame. *istB* (frame -2; 723 bp) of IS*1331A* is located in a relative reading phase of -1 compared with *istA* (frame -1; 1506 bp), in such a way that an ATG initiation codon of *istB* overlaps the TGA stop codon of *istA* by 14 bp (Fig. 6). Also one of the IS*1331* copies has the 5 bp deletion in the *istA* gene (Fig. 3C).

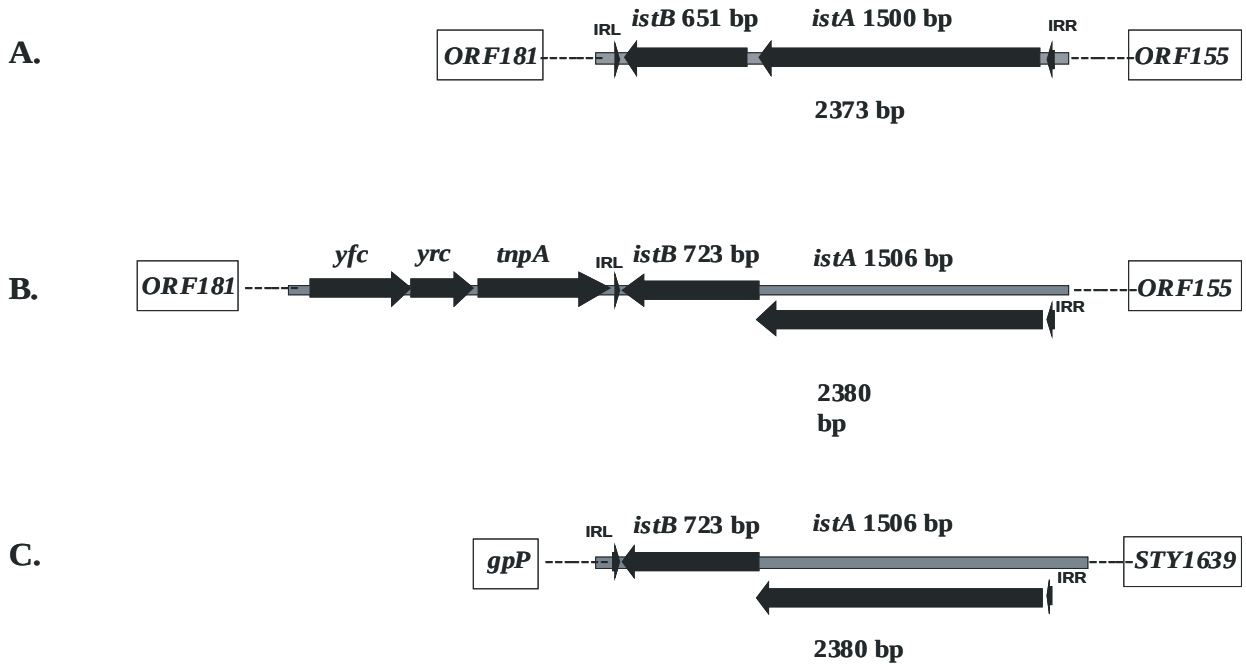


Fig. 2: Schematic structure of the three copies of IS*1331*. IRL – left inverted repeat; IRR – right inverted repeat. Partially sequenced up- and downstream genes with respect to IS*1331* are boxed. The size of the each copy of IS*1331* (including IRL and IRR) is shown in bp under each corresponding structure. A. IS*1331B* on pYV-like plasmid from *Y. enterocolitica* Y11 O:3. B. IS*1331A* on pYV-like plasmid from *Y. enterocolitica* JD E029 O:1. C. IS*1331A* on the chromosome of *Y. enterocolitica* Y11 O:3

A. IRL

```
ORF181  - tgtcaattatggtctaaaagt
tnpA*-  - tgtcaattatggtctaaaagt
ycdX -  - tgtcaattatggtctaaaagt
gpP  -  - tgtcaattatggtctaaaagt
?**- -  - tgtcaattatggtctaaaagt
```

B. IRR

```
actttgcacaatatttaaca - ORF155
actttgcacaatatttaaca - STY1639
actttgcacaatatttaaca- intU
actttgcacaatatttaaca- ORF19
actttgcacaatatttaaca- ORF155*
actttgcacaatatttaaca- STY4827**
```

C.

```
1. atgcccatcaaaacag------gataaggagttgatttgg ORF181-IS1331B-ORF155
2. atgcccatcaaaacag------gataaggagttgatttgg ycdX  -IS1331B-ORF19
3. atgcccatcaaaacagaataaggataaggagttgatttgg gpP   -IS1331A-STY1639
4. atgcccatcaaaacaggataag-----ggagttgatttgg ?     -IS1331A-intU
5. atgcccatcaaaacagaataaggataaggagttgatttgg ?     -IS1331A-STY4827**
6. atgcccatcaaaacagaataaggataaggagttgatttgg tnpA  -IS1331A-ORF155*
```

Fig. 3: A & B. Nucleotide sequences of the IRL and IRR, respectively. Diverse nucleotides in inverted repeats are boxed. The closest ORFs are shown besides of the nucleotide sequences. **C. Partial nucleotide sequence comparison of the part of *istA* from all IS*1331* copies.** The 6-bp and 5-bp deletions are shown as the gaps. From the right side of each sequence the corresponding copy of IS*1331* is exposed.

* - *Y. enterocolitica* JD E029 O:1

** - *Y. enterocolitica* gk1142 O:2

? - downstream gene is unknown

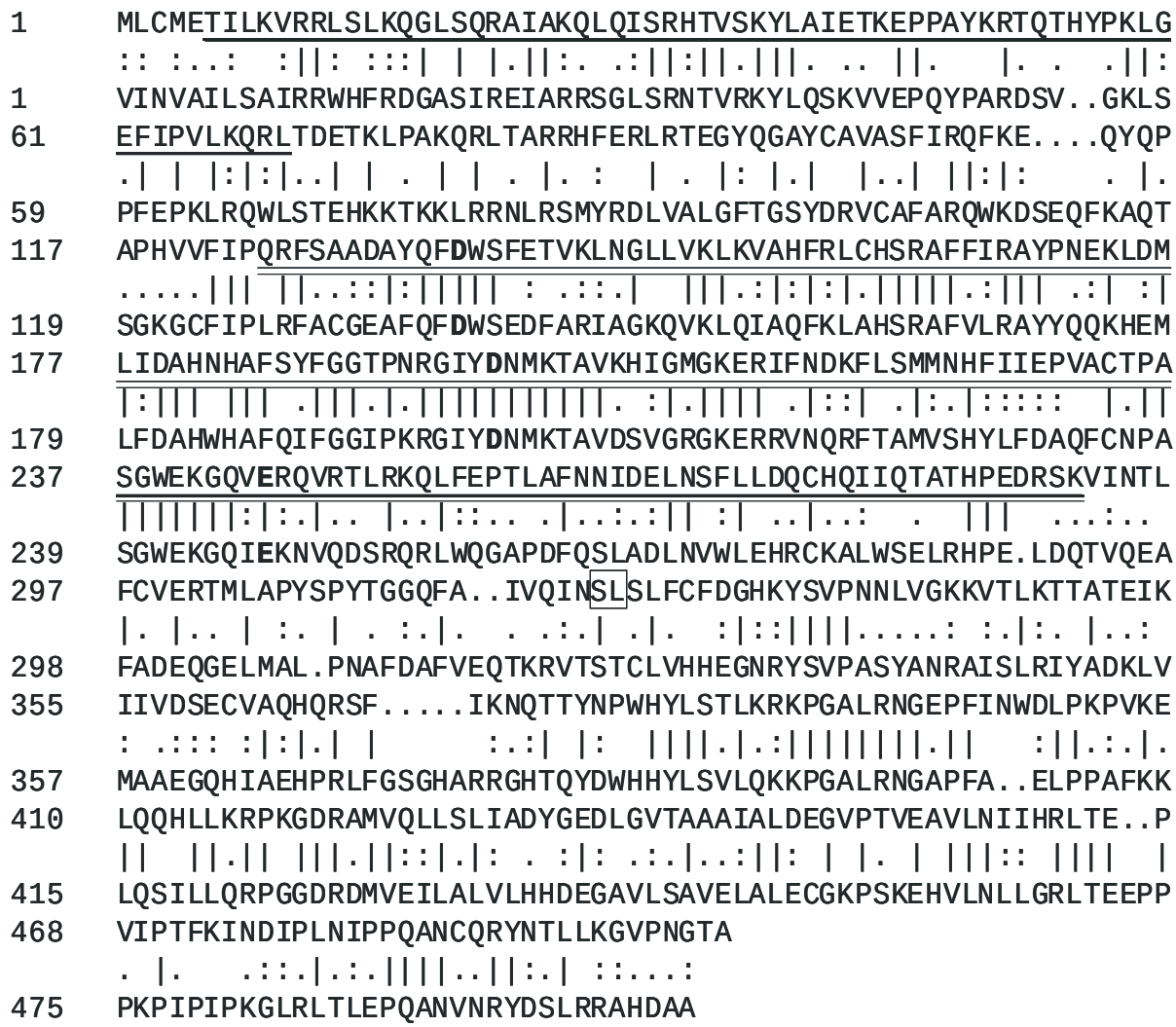


Fig. 4: Amino acid sequence comparison of the IstA transposases from IS*1326* (lower line) and IS*1331A* (upper line). Identical residues are shown with vertical lines, and residues homologous to those in IS*1331A* are pointed or double pointed. The DD(45)E transposase motif in bold; HTH, helix-turn-helix motif, is underlining; an integrase-like core domain in IstA is double underlining. Two absent in IS*1331B* amino acids are boxed.

```
Query: 1   MERHECIEILKQLKLTAMAENFDDVVIDGIRRKRSTMDIIGNLLTTEQTQRHIRSIGYRI
           M+    + ILK LKL  MA   +++           + ++ +L+  E  +R +RS+ Y++
Sbjct: 2   MQHEGHVRILKSLKLFGMAHAIEELGNQNSPAFNQALPMLDSLIKAEVAEREVRSVNYQL
Query: 61  NQARFPQHKTLSDFEFEQSPLNKPSIELLNDCDYIREKRNIIFVGGPGTGKTHLATALGI
             A+FP ++ L  F+F QS +N+ +++ L+ CD++ + +N++ +GGPGTGKTHLATA+G
Sbjct: 62  RVAKFPVYRDLVGFDFSQSLVNEATVKQLHRCDFMEQAQNVVLIGGPGTGKTHLATAIGT
Query: 121 NAATN-GFKVRFWNVLDLVNKLELDKES-KQFKLTNQLTKLDLIVLDDLGYLPFSQKGGA
            A  +   +VRF++ +DLVN LE +K S +Q ++ N+L   DL++LD+LGYLPFSQ GGA
Sbjct: 122 QAVMHLNRRVRFFSTVDLVNALEQEKSSGRQGQIANRLLYADLVILDELGYLPFSQTGGA
Query: 179 LLFHLISQLHEHTSIMITTNLAFSEWVKLFADEKMTAALLDRLVHHCDIIETGNESFRFK
           LLFHL+S+L+E TS+++TTNL+FSEW ++F DEKMT ALLDRL HHC I+ETGNES+RFK
Sbjct: 182 LLFHLLSKLYEKTSVILTTNLSFSEWSRVFGDEKMTTALLDRLTHHCHILETGNESYRFK
Query: 239 NRS
           + S
Sbjct: 242 HSS
```

Fig. 5: Amino acid sequence comparison of the IstB NTP-binding proteins from IS*1326* (lower line) and IS*1331A* (upper line). Identical residues are shown as a consensus residues between two sequences, residues homologous to those in IS*1331A* are shown with plus. The NTP-binding P-loop motif in IstB are indicated by underlining.

In contrast to *istB* of IS*1331A*, *istB* of IS*1331B* (651 bp) is in the same relative phase -2 like *istA* frame, but divided from *istA* by 57 bp intergenic region (Fig. 2). IstA of IS*1331* has a usual for the IS*21* family DD(46)E transposase motif, a putative DNA-binding domain (domain with helix-turn-helix structural motif, 6-70 aa) in the N-terminal part and an integrase-like core domain in the middle of the protein (125-291 aa) (Fig. 4). The 6 bp deletion in *istA* of IS*1331B*, which results in deletion of two amino acids (S and L) in transposase IstA, probably did not affect the important domains in IstA (Fig. 4). IstB contains an ATP/GTP binding P-loop motif (100-118 aa) (Fig. 5).

Based on the similarity to the NTP-binding protein, transposase and presence of the inverted repeats of 30 bp IS*1331* might be classified as a member of the IS*21* family. The transcription

of both *istA* and *istB* was confirmed in *Y. enterocolitica* Y-108C by the reverse transcription PCR (data not shown).

Fig. 6: Partial nucleotide sequence of IS*1331A*. IRL and IRR are shown by arrows. -35 and -10 regions of a putative promoter of *istAB* are underlined. Ribosome binding sites (RBS) are over lined. Start codons of IstA and IstB are boxed, stop codons are shown with asterisks.

2.2 Determination of the copy number and flanking sequences

The Southern blot analysis was performed to determine the copy number of IS*1331*. Hybridization experiments with the 427-bp PCR fragment obtained with the primers

istAM.for and istAM.rev, internal to the IS*1331*, were carried out. For this, genomic DNA of different yersiniae was digested with *Eco*RI (which does not cut inside the IS*1331*), transferred to a nylon membrane and hybridized with the DIG-labeled 427-bp probe (see *Materials and Methods*).

In total, twenty six strains were studied by the Southern blot analysis (Fig. 7).

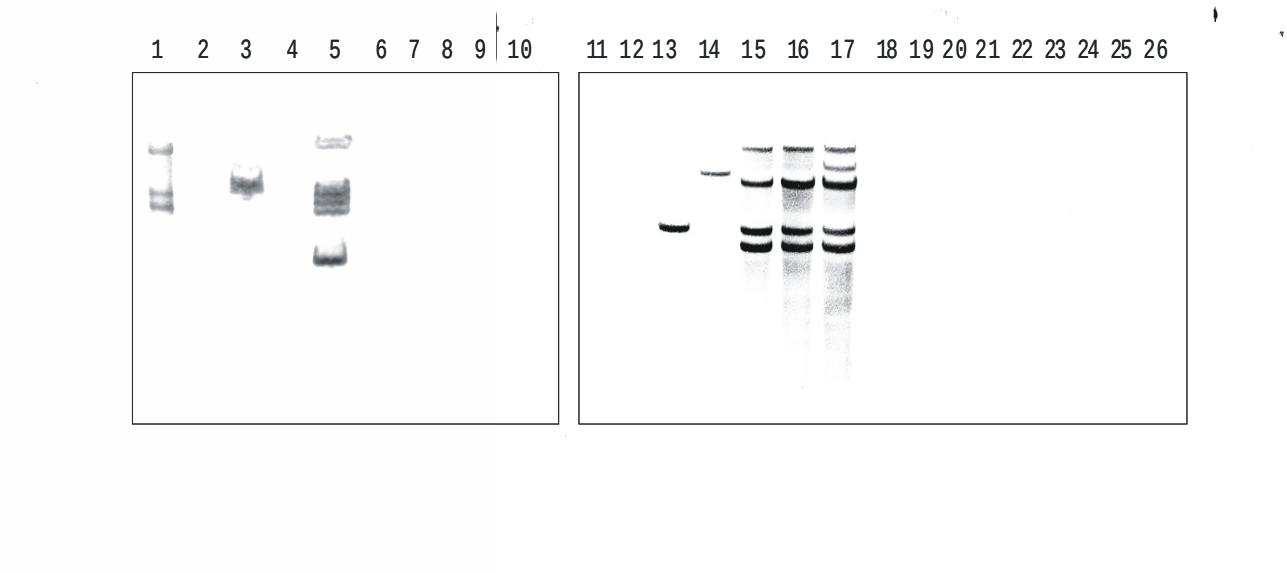


Fig. 7. Southern blot hybridization of *Eco*RI-digested bacterial genomic DNA with a DIG-labeled probe to IS*1331*.

1. *Y. enterocolitica* Y-108C ST O:3 BG 4
2. *Y. enterocolitica* NFO ST O:5 BG 1A
3. *Y. enterocolitica* ST O:5,27 BG 2
4. *Y. enterocolitica* WA-C ST O:8 BG 1B
5. *Y. enterocolitica* 96C ST O:9 BG 2
6. *Y. pstbc*. PBI ST O:1A
7. *Y. pstbc*. 346 ST O:3
8. *Y. pestis* EV
9. *Y. pestis* G8786
10. *Y. pestis* Yokohama
11. *Y. enterocolitica* 438/80 ST O:6,31
12. *Y. enterocolitica* 189/80 ST O:6,30 BG 1A
13. *Y. enterocolitica* gk 1142 ST O:2
14. *Y. enterocolitica* JD E029 ST O:1
15. *Y. enterocolitica* 221 Erg. Nod ST O:3 BG 4
16. *Y. enterocolitica* 2118-Y ST O:3 BG 4
17. *Y. enterocolitica* S-2840 ST O:3 BG 4
18. *Y. enterocolitica* ST O:36 BG 1A
19. *Y. ruckeri* 529-36/85
20. *Y. mollaretii* ST O:59 H279-36/86
21. *Y. bercovieri* ST O:16 H632-36/85
22. *Y. rohdei* ST O:76 H274-36/78
23. *Y. aldovae* ST O:NT H344-36/91
24. *Y. kristensenii* ST O:50 H25-36/84
25. *Y. frederiksenii* ST O:60 H56-36/81
26. *Y. intermedia* ST O:17 H9-36/83

Eight tested strains possess the IS element. The number of IS*1331* copies in the genome varied among the strains. From the hybridisation patterns we assumed that *Y. enterocolitica* O:3 221 Erg. Nod., 2118-Y and S-2840 (all pYV-bearing strains) contain at least four IS*1331* copies (Fig. 7), in comparison to three positive bands, which were detected in the plasmid-cured derivative of *Y. enterocolitica* Y-108C. It indicates that at least one copy of IS*1331* is located on the pYV-like plasmid. DNA isolated from other low pathogenic *Y. enterocolitica* H567/90 ST O:5,27 and *Y. enterocolitica* 96C ST O:9 gave only one and five positive signals under stringent condition, respectively. It is interesting, that *Y. enterocolitica* S-2840 ST O:3 has five positive signals instead of four detected in other strains of ST O:3. *Y. enterocolitica* gk 1142 ST O:2 and *Y. enterocolitica* JD E029 ST O:1 contain at least one copy of IS*1331* (Fig. 7).

All accessible nucleotide sequences flanking IS*1331* in *Y. enterocolitica* Y11 ST O:3 strain were obtained from the draft of the genome sequencing project of the above-mentioned strain (Table 9). We have additionally performed sequencing beside IS*1331* insertion sites in another *Y. enterocolitica* Y-108P ST O:3 with the original Y11 PCR primers. It was not surprising that IS*1331* insertion sites in *Y. enterocolitica* Y-108P were the same like in Y11. We have also determined genomic sequences flanking IS*1331* in *Y. enterocolitica* JD E029 ST O:1 and *Y. enterocolitica* gk 1142 ST O:2 by inverse PCR. Afterwards, the original Y11 PCR primers were used to verify the sequencing data. As expected, IS*1331* locations in *Y. enterocolitica* JD E029 ST O:1 and *Y. enterocolitica* gk 1142 ST O:2 were different. IS*1331* in *Y. enterocolitica* JD E029 ST O:1 strain was found on the pYV-like plasmid in the ORF181-ORF155 intergenic region (Fig. 2), while in *Y. enterocolitica* gk 1142 ST O:2 IS*1331* is chromosomally located in the neighbourhood of the putative ORF with similarity to STY4827 (putative phage capsid protein) of *Salmonella enterica* (Acc. NP_458905). Surprisingly, ORF181-ORF155 intergenic region of pYV-like plasmid of *Y. enterocolitica* JD E029 O:1 contains three additional ORFs – *yfc*, *yrc* and *tnpA* besides IS*1331A* (Fig. 2). Yfc has the similarity to Plu4880 (74% at amino acid level, encodes the uncharacterized conserved protein from *Photorhabdus luminescens* subsp. *laumondii* TTO1, Acc. CAE17252). Yrc has the similarity to Raeut03006195 (88% at amino acid level), which encodes the putative site-specific recombinase from *Ralstonia eutropha* JMP134 (Acc. ZP_00165678). *tnpA* is a pseudogene with the similarity to the *tnpA* from pKLH466 plasmid (*Pseudomonas* sp. LS46-6, Acc. CAC80084).

Table 9. IS*1331* flanking ORFs in the *Y. enterocolitica* Y11 and Y-108P ST O:3 genomes.

No	Downstream gene	Similar gene/ product/organism	Similarity (positives amino acids, %)	Acc. No.***	Upstream gene	Similar gene/ product/organism	Similarity (positives amino acids, %)	Acc. No.***
1	ORF181	ORF181/unknown/ *Yersinia enterocolitica* W22703 O:9 pYVe227 plasmid	159/160 (99%)	AAD16801	ORF155	ORF155/ unknown/ *Yersinia enterocolitica* W22703 O:9 pYVe227 plasmid	154/155 (99%)	AAD16802
2	*ycdX*	ycdX/ hypothetical protein ycdX precursor/ *Escherichia coli* CFT073	58/89 (65%)	NP_753209	ORF19	ORF19 (TraC-3)/ antirestriction protein/ *Escherichia coli* ECOR31	230/252 (91%)	AAP70295
3	*gpP*	gpP/ terminase, DNA-dependent ATPase/ Enterobacteria phage P2	178/202 (88%)	AAD03269	STY1639	STY1639/ putative bacteriophage tail fiber assembly protein/*Salmonella enterica* serovar Typhi CT18	64/106 (60%)	NP_456050
4	unknown				*intU*	intU/ putative integrase for prophage CP-933U/ *Shigella flexneri* 2a str. 301	155/313 (49%)	NP_707880
5*	unknown				STY4827	STY4827/putative phage capsid protein/*Salmonella enterica* serovar Typhi	51/75 (68%)	NP_458905
6**	*tnpA*	Transposase/pKLH466/ *Pseudomonas* sp. LS46-6	130/148 (87%)	CAC80084	ORF155	ORF155/ unknown/ *Yersinia enterocolitica* W22703 O:9 pYVe227 plasmid	40/40 (100%)	AF102990

* Located on the chromosome of *Y. enterocolitica* O:2
** Located on the pYV-like plasmid of *Y. enterocolitica* O:1
*** Accession number for homological proteins

2.3 Distribution of IS*1331* among various yersiniae

The distribution of the newly described IS*1331* element in the genomes of the pathogenic and non-pathogenic yersiniae species *Y. pestis*, *Y. pseudotuberculosis*, *Y. enterocolitica*, *Y. ruckeri*, *Y. mollaretii*, *Y. bercovieri*, *Y. rohdei*, *Y. kristensenii*, *Y. frederiksenii* and *Y. intermedia* was investigated by Dot blot analysis. Hybridization experiments with the above mentioned DIG-labeled PCR-fragment, which is internal to the IS*1331*, were carried out. Surprisingly, hybridization signals were obtained only with the genomic DNA of the weakly pathogenic *Y. enterocolitica* bio- and serotypes (Table 10). IS*1331* was not identified in any other yersiniae tested under stringent conditions.

Table 10. Strains used in the screening of IS*1331* .

No.	Isolate	Origin	Country	IS*1331**
	Y. enterocolitica ST O:3 BG 4			
1	Y-108C	Man	Germany	+
2	Y11	Man	Germany	+
3	5307	Man	unknown	+
4	5783	Man	unknown	+
5	Y486	Calf	unknown	+
6	Y626 WS 15/93	Pig	Belgium	+
7	Y633 (WE50/01)	Man	Belgium	+
8	Y641	Pup	unknown	+
9	560 (SW13123)	Pig	Japan	+
10	559 (M388)	Pig	Japan	+
11	556 (8265)	Man	France	+
12	531 (SW13711)	Pig	Japan	+
13	554 (29C-43a)	Man	Norway	+
14	555 (MCH697)	Man	Canada	+
15	Y485	Pig	unknown	+
16	628 W 498	Pig tongues	Belgium	+
17	642	unknown	Germany	+
18	Y631	Pig/food technology	Denmark	+
19	S-2840	Man	Germany	+
20	2118-Y	Man	Germany	+
21	221 Erg.Nod	Man	Germany	+
22	Y-1088	Man	Germany	+
23	66 Stool	Man	Germany	+
24	56 LK	Man	Germany	+
25	62 Stool	Man	Germany	+
26	59 LK	Man	Germany	+
27	4147 pYVII Ar+	Man	Germany	+
28	7347 pYVI Ar+	Man	Germany	+
29	Y745 IP24231	Man	Japan	+
30	Y746 IP24232	Man	Japan	+
31	Y747 IP134	Man	Sweden	+
32	Y748 IP21981	Man	France	+
33	Y749 IP1601	Man	Japan	+
34	Y750 IP19718	Man	China	+
35	Y751 IP23222	Man	UK	+
36	Y752 IP23357	Man	Brazil	+

Table 10. Continued

No.	Isolate	Origin	Country	IS*1331**
37	Y753 IP24309	Man	N. Caledonia	+
38	Y754 IP24310	Man	N. Caledonia	+
39	Y755 IP7032	Man	South Africa	+
40	Y756 IP7036	Man	South Africa	+
41	Y757 IP3692	Man	Hungary	+
42	Y758 IP3704	Man	Hungary	+
43	Y759 IP4115	Man	Canada	+
44	Y763 IP4124	Man	Canada	+
45	Y764 IP4125	Man	Canada	+
46	Y765 IP22274	Man	Australia	+
47	Y766 IP22275	Man	Australia	+
48	Y767 IP22276	Man	Australia	+
49	Y768 IP25728	Man	Australia	+
50	Y769 IP23227	Man	New Zealand	+
51	Y770 IP23228	Man	New Zealand	+
52	Y771 IP23230	Man	New Zealand	+
53	910/98	Man	Poland	+
54	562/03	Man	Poland	+
55	465/00	Man	Poland	+
56	128/99	Man	Poland	+
57	627/99	Man	Poland	+
58	120/98	Man	Poland	+
59	243/96	Man	Poland	+
60	504/97	Man	Poland	+
61	99/96	Man	Poland	+
62	457/98	Man	Poland	+
63	567/98	Man	Poland	+
64	265/01	Man	Poland	+
65	274/99	Man	Poland	+
66	188/01	Man	Poland	+
67	254/97	Man	Poland	+
68	353/96	Man	Poland	+
69	332/98	Man	Poland	+
70	184/97	Man	Poland	+
	Y. enterocolitica ST O:9 BG 2			
71	96C (Y-96)	Man	Germany	+
72	564 (7Oulua)	Man	Finland	+
73	563 (YE099)	Man	Canada	+
74	Y738 IP22393	Man	France	+

Table 10. Continued

No.	Isolate	Origin	Country	IS*1331**
	Y. enterocolitica ST O:5,27 BG 2			
75	586 (H567/90)	Man	Germany	+
76	568 (YE873)	Pig	Canada	+
77	534 (D113)	Dog	Japan	+
78	Y740 IP199	Man	USA	+
79	Y741 IP885	Man	UK	+
80	Y743 IP22460	Man	Australia	+
	Y. enterocolitica ST O:1			
81	JD E029	human	unknown	+
	Y. enterocolitica ST O:2			
82	gk1142	hare	unknown	+
	Y. enterocolitica ST O:1,2,3 BG 3			
83	JD E766	human	unknown	+
	Y. enterocolitica ST O:2a,2b,3			
84	Y772 IP1	Hare	France	-
85	Y773 IP178	Hare	France	-
	Y. enterocolitica ST O:6,31			
86	438/80	unknown	unknown	-
87	1309/80	unknown	unknown	-
	Y. enterocolitica ST O:36 BG 1A			
88	IP 2222	Nonclinical isolate	unknown	-
	Y. enterocolitica ST O:4			
89	gc 3973-76	Man	USA	-
	Y. enterocolitica ST O:4,32			
90	JD E701	human	unknown	-
	Y. enterocolitica ST O:5 BG 1A			
91	NF-O	Man	Newfoundland	-
92	Y755 IP124	Pony	France	-
	Y. enterocolitica ST O:6,30-6,61			
93	Y774 IP102	unknown	Denmark	-
	Y. enterocolitica ST O:6,30 BG 1A			
94	189/80	unknown	unknown	-
95	6737/80	unknown	unknown	-
	Y. enterocolitica ST O:7,8 BG 1A			
96	350	unknown	Germany	-
	Y. enterocolitica ST O13:b,e,f,i			
97	329/82	unknown	USA	-
	Y. enterocolitica ST 020:b,e,f,i			

Table 10. Continued

No.	Isolate	Origin	Country	IS*1331**
98	3039/85	unknown	unknown	-
	Y. enterocolitica ST 021:b,e,f,i			
99	209/84	unknown	unknown	-
	Y. enterocolitica ST O:41,43			
100	478/91	unknown	unknown	-
	Y. enterocolitica ST O:48			
101	4643/93	unknown	unknown	-
	Y. enterocolitica ST O:8 BG 1B			
102	WA-C (WA-314)	Man	USA	-
103	589 (900/90)	Man	Japan	-
104	575 (893/87)	Man	Italy	-
105	Y736 IP636	unknown	USA	-
	Y. pseudotuberculosis			
106	PB1, ST O:1	unknown	unknown	-
107	735	Rattus norvegicus	St. Petersburg, Russia	-
108	952	Rattus norvegicus	St. Petersburg, Russia	-
109	346, ST O:3	Man	Denmark	-
	Y. pestis			
110	5764	Rattus norvegicus	St. Petersburg, Russia	-
111	G8786	Vole	Georgia	-
112	EV	Man	Madagascar	-
113	Yokohama	Man	Japan	
	Other species			
114	*Y. intermedia* H9-36/83 ST O:17	unknown	unknown	-
115	*Y. frederiksenii* H56-36/81 ST O:60	unknown	unknown	-
116	*Y. kristensenii* H25-36/84 ST O:50	unknown	unknown	-
117	*Y. rohdei* H274-36/78 ST O:76	unknown	unknown	-
118	*Y. bercovieri* H632-36/85 ST O:16	unknown	unknown	-
119	*Y. mollaretii* H279-36/86 ST O:59	unknown	unknown	-
120	*Y. ruckeri* 529-36/85	unknown	unknown	-
121	*Y. aldovae* H344-36/91 ST O:NT	unknown	unknown	-

MvP – Max von Pettenkofer Institute of Hygiene and Medical Microbiology, Munich, Germany
PZH - Department of Bacteriology National Institute of Hygiene, Warsaw, Poland
SaBw - Institute for Microbiology, German Federal Armed Forces, Munich, Germany
SESAHS - South Eastern Sydney Area Health Service, Sydney, Australia
*„+“ – IS*1331* is present in a genome; „-“ – IS*1331* is absent in a genome

3. Uncovering a novel RTX-like toxin in *Y. enterocolitica* subsp. *palearctica* Y-108

One of the subtracted fragments (01_52), uncovered by SSH, showed high similarity to the *rtxA* gene of *Vibrio cholerae* El Tor. To obtain the complete sequence of the putative gene we screened the cosmid gene bank of *Y. enterocolitica* Y-108C. Two cosmids (pSC175D and pSC1012B) carrying the ORF of interest were isolated and subsequently sequenced. A genetic locus comprising 14665 bp and containing 7 open reading frames was uncovered (Fig. 8).

These ORFs have been designated accordingly to the similarity of their predicted products (Table 11).

Figure 9 compares the organization of the RTX cluster with the homologous that encoded by *E. coli*, *V. cholerae*, *Photorhabdus luminescens* and sequence of the neighbor genes from from the *Y. pestis* genome.

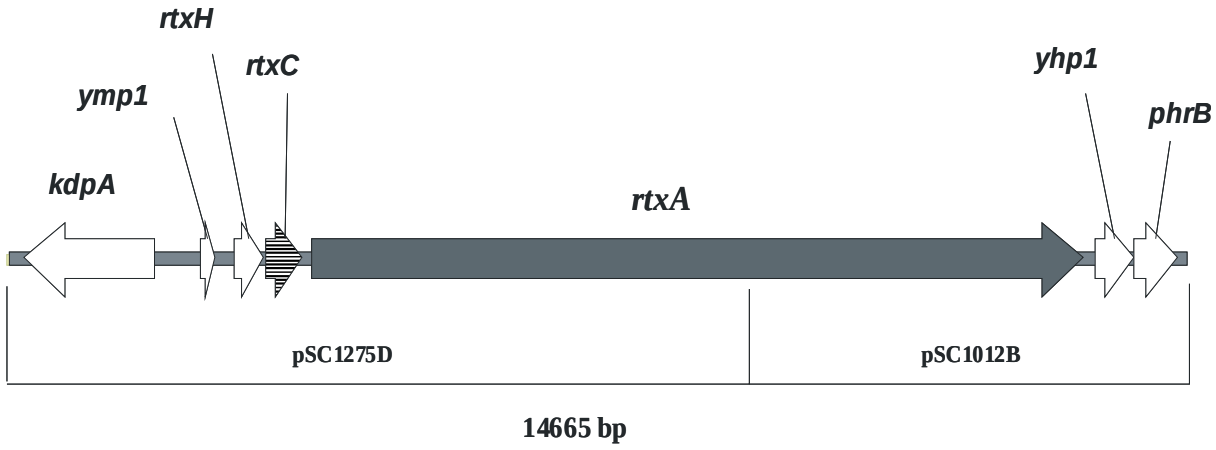


Fig. 8: Organization of the RTX-gene cluster and neighbouring ORFs
ORFs are shown in italic; pSC1275D and pSC1012B illustrate, which part of the *rtxA* cluster comes from the sequenced cosmids.

Toxin Activator protein Hypothetical protein

Table 11. Background and RTX-gene cluster of *Y. enterocolitica* Y-108C

ORF/size	Similar gene/Putative product/Microorganism	Homology, aa
kdpA 1638 bp	*kdpA*/putative potassium-transporting ATPase A chain/*Yersinia pestis* CO92	470/524 (89%)
ymp1 204 bp	YPO2693/putative membrane protein/*Yersinia pestis* CO92	61/68 (89%)
rtxH 372 bp	VC1449/hypothetical protein/*Vibrio cholerae* O1 El Tor	79/107 (73%)
rtxC 456 bp	VC1450/cytolysin-activating lysine-acyltransferase/*Vibrio cholerae* O1 El Tor	111/151 (73%)
rtxA 9636 bp	VC1451/RTX toxin/ *Vibrio cholerae* O1 El Tor	1972/3212 (61%)
yhp1 474 bp	YPO2694/hypothetical protein/ *Yersinia pestis* CO92	126/149 (84%)
phrB 546 bp	YPO2695/putative deoxyribodipyrimidine photolyase/ *Yersinia pestis* CO92	141/178 (79%)

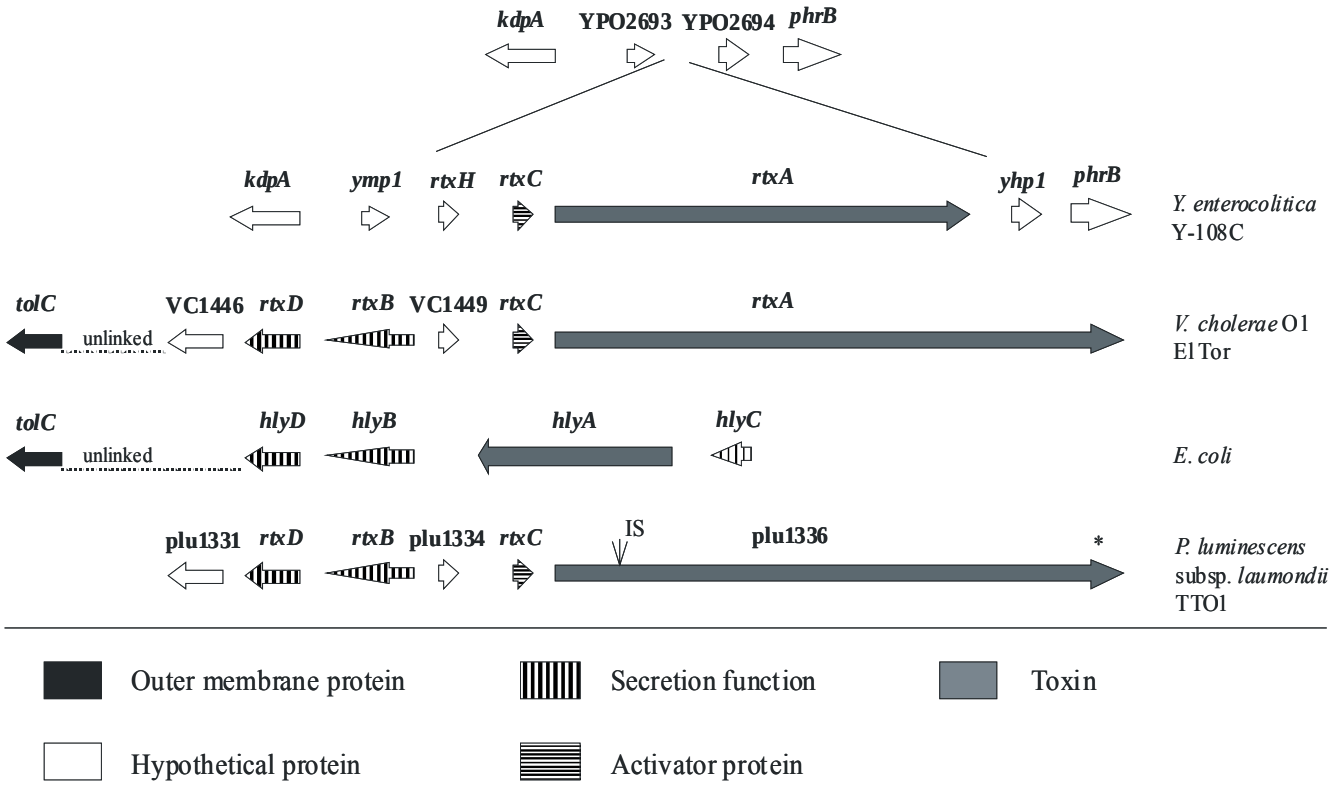


Fig. 9: Genomic structure of the RTX cluster in *Y. enterocolitica* Y-108C; line-up of the RTX gene cluster and neighbouring genes with the RTX elements from *V. cholerae*, *E. coli* and *P. luminescens*.

The known or predicted functions of each ORF are indicated by different shadings. * - The *rtxA* homologs from *P. luminescens* are clustered in two chromosomal regions and tandemly organized. Four of them are complete genes and other four were disrupted by frameshifts or insertion sequences. ↓IS – indicates insertion sequence element.

3.1 Structure of the *rtx* operon

We have identified and characterized the gene cluster in *Y. enterocolitica* Y-108C ST O:3 that contains four genes: *rtxA*, *rtxC*, *ymp1*, and *rtxH* (Fig. 8, 9). The possible toxin, RtxA, resembles members of the RTX (repeats in toxin) toxin family that contains a glycin-rich repeated motif. Like other RTX toxins, it is associated with an activator, RtxC – acyltransferase. In the case of the cholerae's cytotoxin and most other RTX toxins, the genes required for toxin biosynthesis and secretion exist in one operon (Fig. 9). Within the operon should be two genes encoding secretion proteins that are components of ABC transporter system (*rtxB* and *rtxD*). In the *Y. enterocolitica* Y-108C genome we have not found yet transporter genes which might transport RtxA. But we have found two ORFs with unclear functions. One of them, Ymp1, has similarity to a putative membrane protein YPO2693 from *Y. pestis* CO92 and second one, RtxH, has similarity to a peptide chain release factor 1 (VC1449) from *V. cholerae*.

Comparative analysis (Fig. 10) of the RtxA from *Y. enterocolitica* and *V. cholerae* as well as the RtxC proteins on amino acid level showed that these proteins have more similarity to each other than to the other proteins of RTX-family, for instance, HlyA and HlyC from *E. coli*. (Fig. 11)

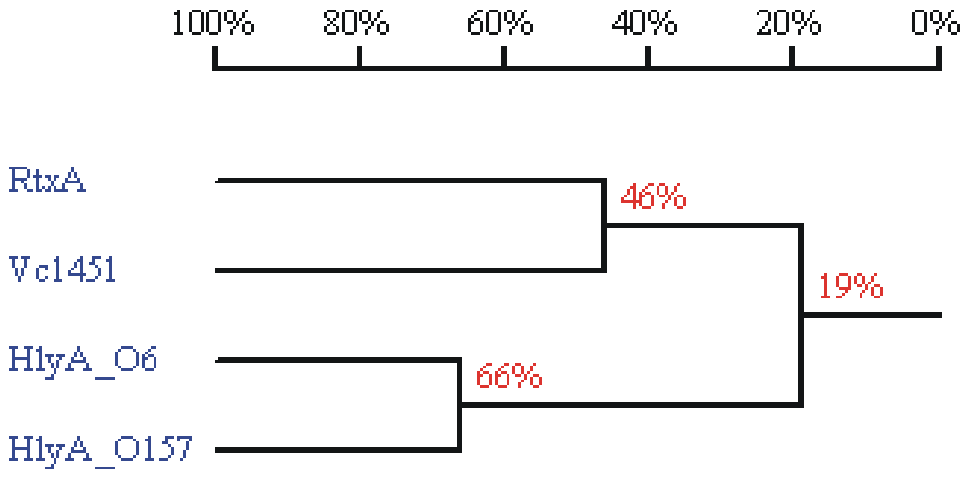


Fig. 10: Homology tree of RtxA-like proteins. RtxA – RtxA-like protein of *Y. enterocolitica* Y11; Vc1451 – RtxA protein of *V. cholerae*; HlyA_O6 – HlyA protein of *E. coli* O:6; HlyA_O157 – HlyA protein of *E. coli* O:157. Numbers show the level of similarity.

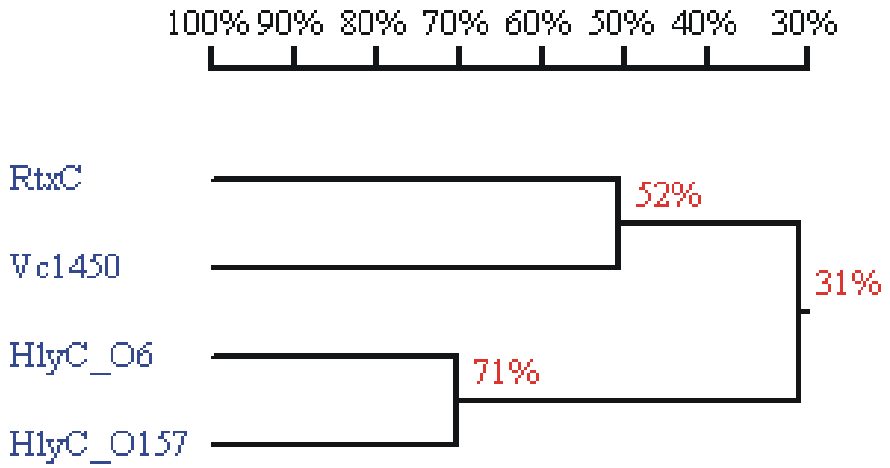


Fig. 11: Homology tree of RtxC-like proteins. RtxC – RtxC-like protein of *Y. enterocolitica* Y11; Vc1450 – RtxC protein of *V. cholerae*; HlyC_O6 – HlyC protein of *E. coli* O:6; HlyC_O157 – HlyC protein of *E. coli* O:157. Numbers show the level of similarity.

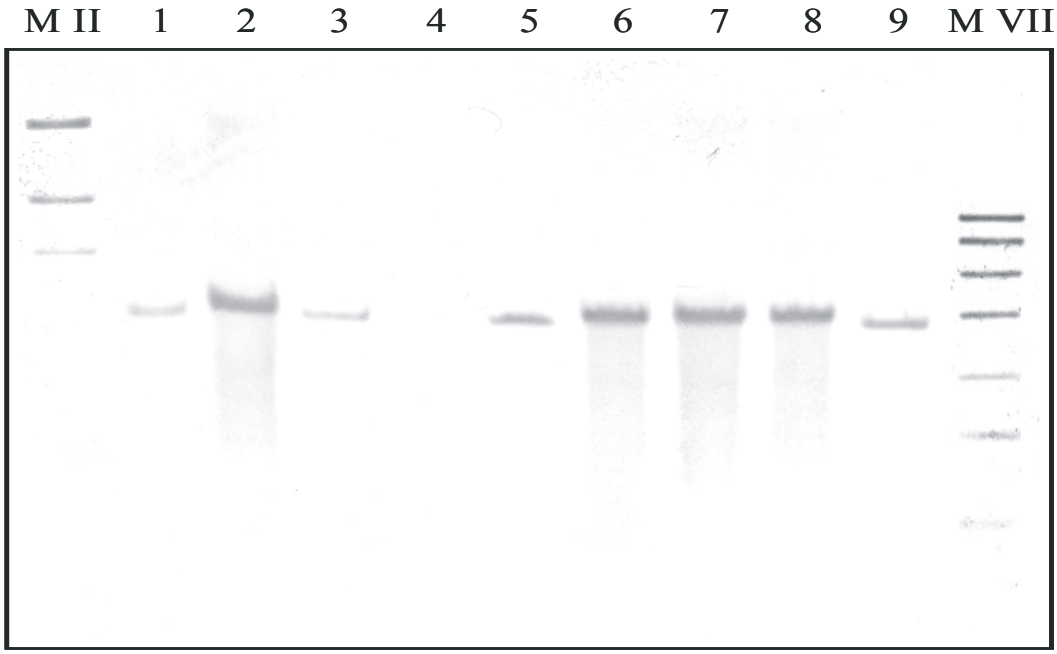


Fig. 12: Southern blot hybridization of *Hind*III-digested genomic DNA of various *Yersinia* strains with DIG-labeled probe to *rtxA*.
M II. DIG-labeled DNA molecular weight marker II, 23130, 9416 and 6557 bp fragments;
1. *Y. enterocolitica* JD E029 ST O:1;
2. *Y. enterocolitica* 2118-Y ST O:3 BG 4;
3. *Y. enterocolitica* Y11 ST O:3 BG 4;
4. *Y. frederiksenii* H-56-36/81 ST O:60;
5. *Y. enterocolitica* BV4- ST O:3 BG 4;
6. *Y. enterocolitica* gk 1142 ST O:2;
7. *Y. enterocolitica* 4147 pYVII Ar+ ST O:3 BG 4;
8. *Y. enterocolitica* 5307 Ar+ O:3 BG 4;
9. *Y. enterocolitica* 5783 Ar- ST O:3 BG 4;
M VII. DIG-labeled DNA molecular weight marker VII, 8576, 7427, 6106, 4899, 3639, 2799 and 1953 bp fragments.

3.2 Distribution of the RTX-like cluster among different *Yersinia*

We have established by using the Southern blot analysis and DNA microarray that *rtxA* is present only among weakly pathogenic *Y. enterocolitica* strains and is absent in highly and nonpathogenic bioserotypes (Table 12 and Fig. 12).

Table 12. Distribution *rtxA* among different yersiniae.

No.	Isolate	Origin	Country	*rtxA**
	Y. enterocolitica ST O:3 BG 4			
1	Y-108C	Man	Germany	+
2	Y11	Man	Germany	+
3	5307	Man	unknown	+
4	5783	Man	unknown	+
5	Y486	Calf	unknown	+
6	Y626 WS 15/93	Pig	Belgium	+
7	Y633 (WE50/01)	Man	Belgium	+
8	Y641	Pup	unknown	+
9	560 (SW13123)	Pig	Japan	+
10	559 (M388)	Pig	Japan	+
11	556 (8265)	Man	France	+
12	531 (SW13711)	Pig	Japan	+
13	554 (29C-43a)	Man	Norway	+
14	555 (MCH697)	Man	Canada	+
15	Y485	Pig	unknown	+
16	628 W 498	Pig tongues	Belgium	+
17	642	unknown	Germany	+
18	Y631	Pig/food technology	Denmark	+
19	S-2840	Man	Germany	+
20	2118-Y	Man	Germany	+
21	221 Erg.Nod.	Man	Germany	+
22	Y-1088	Man	Germany	+
23	66 Stool	Man	Germany	+
24	56 LK	Man	Germany	+
25	62 Stool	Man	Germany	+
26	59 LK	Man	Germany	+
27	4147 pYVII Ar+	Man	Germany	+
28	7347 pYVI Ar+	Man	Germany	+
29	Y745 IP24231	Man	Japan	+
30	Y746 IP24232	Man	Japan	+
31	Y747 IP134	Man	Sweden	+
32	Y748 IP21981	Man	France	+
33	Y749 IP1601	Man	Japan	+
34	Y750 IP19718	Man	China	+
35	Y751 IP23222	Man	UK	+
36	Y752 IP23357	Man	Brazil	+

Table 12. Continued

No.	Isolate	Origin	Country	*rtxA**
37	Y753 IP24309	Man	N. Caledonia	+
38	Y754 IP24310	Man	N. Caledonia	+
39	Y755 IP7032	Man	South Africa	+
40	Y756 IP7036	Man	South Africa	+
41	Y757 IP3692	Man	Hungary	+
42	Y758 IP3704	Man	Hungary	+
43	Y759 IP4115	Man	Canada	+
44	Y763 IP4124	Man	Canada	+
45	Y764 IP4125	Man	Canada	+
46	Y765 IP22274	Man	Australia	+
47	Y766 IP22275	Man	Australia	+
48	Y767 IP22276	Man	Australia	+
49	Y768 IP25728	Man	Australia	+
50	Y769 IP23227	Man	New Zealand	+
51	Y770 IP23228	Man	New Zealand	+
52	Y771 IP23230	Man	New Zealand	+
53	910/98	Man	Poland	+
54	562/03	Man	Poland	+
55	465/00	Man	Poland	+
56	128/99	Man	Poland	+
57	627/99	Man	Poland	+
58	120/98	Man	Poland	+
59	243/96	Man	Poland	+
60	504/97	Man	Poland	+
61	99/96	Man	Poland	+
62	457/98	Man	Poland	+
63	567/98	Man	Poland	+
64	265/01	Man	Poland	+
65	274/99	Man	Poland	+
66	188/01	Man	Poland	+
67	254/97	Man	Poland	+
68	353/96	Man	Poland	+
69	332/98	Man	Poland	+
70	184/97	Man	Poland	+
	Y. enterocolitica ST O:9 BG 2			
71	96C (Y-96)	Man	Germany	-
72	564 (7Oulua)	Man	Finland	-
73	563 (YE099)	Man	Canada	-
74	Y738 IP22393	Man	France	-

Table 12. Continued

No.	Isolate	Origin	Country	*rtxA**
	Y. enterocolitica ST O:5,27 BG 2			
75	586 (H567/90)	Man	Germany	-
76	568 (YE873)	Pig	Canada	-
77	534 (D113)	Dog	Japan	-
78	Y740 IP199	Man	USA	-
79	Y741 IP885	Man	UK	-
80	Y743 IP22460	Man	Australia	-
	Y. enterocolitica ST O:1			
81	JD E029	human	unknown	-
	Y. enterocolitica ST O:2			
82	gk1142	hare	unknown	-
	Y. enterocolitica ST O:1,2,3 BG 3			
83	JD E766	human	unknown	-
	Y. enterocolitica ST O:2a,2b,3			
84	Y772 IP1	Hare	France	-
85	Y773 IP178	Hare	France	-
	Y. enterocolitica ST O:6,31			
86	438/80	unknown	unknown	-
87	1309/80	unknown	unknown	-
	Y. enterocolitica ST O:36 BG 1A			
88	IP 2222	Nonclinical isolate	unknown	-
	Y. enterocolitica ST O:4			
89	gc 3973-76	Man	USA	-
	Y. enterocolitica ST O:4,32			
90	JD E701	human	unknown	-
	Y. enterocolitica ST O:5 BG 1A			
91	NF-O	Man	Newfoundland	-
92	Y755 IP124	Pony	France	-
	Y. enterocolitica ST O:6,30-6,61			
93	Y774 IP102	unknown	Denmark	-
	Y. enterocolitica ST O:6,30 BG 1A			
94	189/80	unknown	unknown	-
95	6737/80	unknown	unknown	-
	Y. enterocolitica ST O:7,8			
96	350	unknown	Germany	-
	Y. enterocolitica ST O13:b,e,f,i			
97	329/82	unknown	USA	-
	Y. enterocolitica ST 020:b,e,f,i			

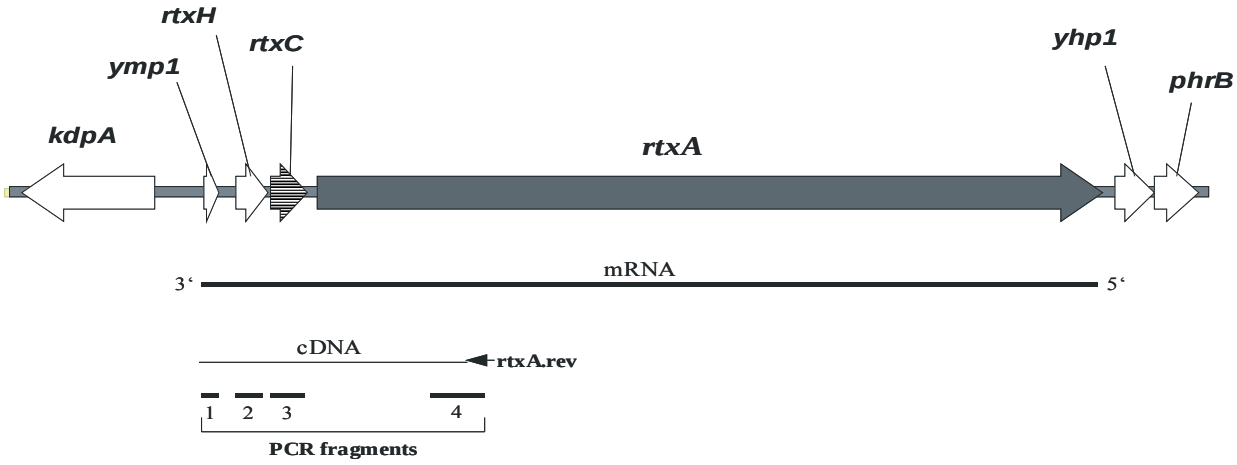
Table 12. Continued

No.	Isolate	Origin	Country	*rtxA**
98	3039/85	unknown	unknown	-
	Y. enterocolitica ST 021:b,e,f,i			
99	209/84	unknown	unknown	-
	Y. enterocolitica ST O:41,43			
100	478/91	unknown	unknown	-
	Y. enterocolitica ST O:48			
101	4643/93	unknown	unknown	-
	Y. enterocolitica ST O:8 BG 1B			
102	WA-C (WA-314)	Man	USA	-
103	589 (900/90)	Man	Japan	-
104	575 (893/87)	Man	Italy	-
105	Y736 IP636	unknown	USA	-
	Y. pseudotuberculosis			
106	PB1, ST O:1	unknown	unknown	-
107	735	Rattus norvegicus	St. Petersburg, Russia	-
108	952	Rattus norvegicus	St. Petersburg, Russia	-
109	346, ST O:3	Man	Denmark	-
	Y. pestis			
110	5764	Rattus norvegicus	St. Petersburg, Russia	-
111	G8786	Vole	Georgia	-
112	EV	Man	Madagascar	-
113	Yokohama	Man	Japan	
	Other species			
114	*Y. intermedia* H9-36/83 ST O:17	unknown	unknown	-
115	*Y. frederiksenii* H56-36/81 ST O:60	unknown	unknown	-
116	*Y. kristensenii* H25-36/84 ST O:50	unknown	unknown	-
117	*Y. rohdei* H274-36/78 ST O:76	unknown	unknown	-
118	*Y. bercovieri* H632-36/85 ST O:16	unknown	unknown	-
119	*Y. mollaretii* H279-36/86 ST O:59	unknown	unknown	-
120	*Y. ruckeri* 529-36/85	unknown	unknown	-
121	*Y. aldovae* H344-36/91 ST O:NT	unknown	unknown	-

*„+“ – *rtxA* is present in a genome; „-“ – *rtxA* is absent in a genome

3.3 Transcription analysis of the RTX gene cluster

Reverse transcription analysis was carried out to determine the transcription of the genes of the RTX cluster. The positive transcripts indicate in vivo transcription of all four ORFs (*ymp1*, *rtxH*, *rtxC*, and *rtxA*) as a single mRNA (Fig. 13).



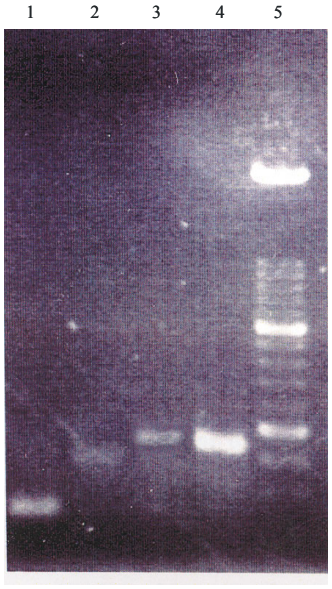


Fig. 13: Gel electrophoresis of the PCR fragments after RT-PCR and schematic representation of the common RTX genes‘ transcription

1. PCR fragment from primers to *ymp1*
2. PCR fragment from primers to *rtxH*
3. PCR fragment from primers to *rtxC*
4. PCR fragment from primers to *rtxA*
5. Molecular weight marker

3.4 Structural features of RtxA

The RTX toxins as a group have certain common domain structures: an N-terminal hydrophobic domain required for pore formation; central prototoxin activation sites; C-terminal glycine-rich (GD) calcium-binding repeats involved in target-cell binding; and a C-terminal signal for secretion.

Analysis of the deduced amino acid sequence of the *Y. enterocolitica* RtxA shows some similar as well as distinguishing features compare to the *V. cholerae* RtxA (Fig. 14).

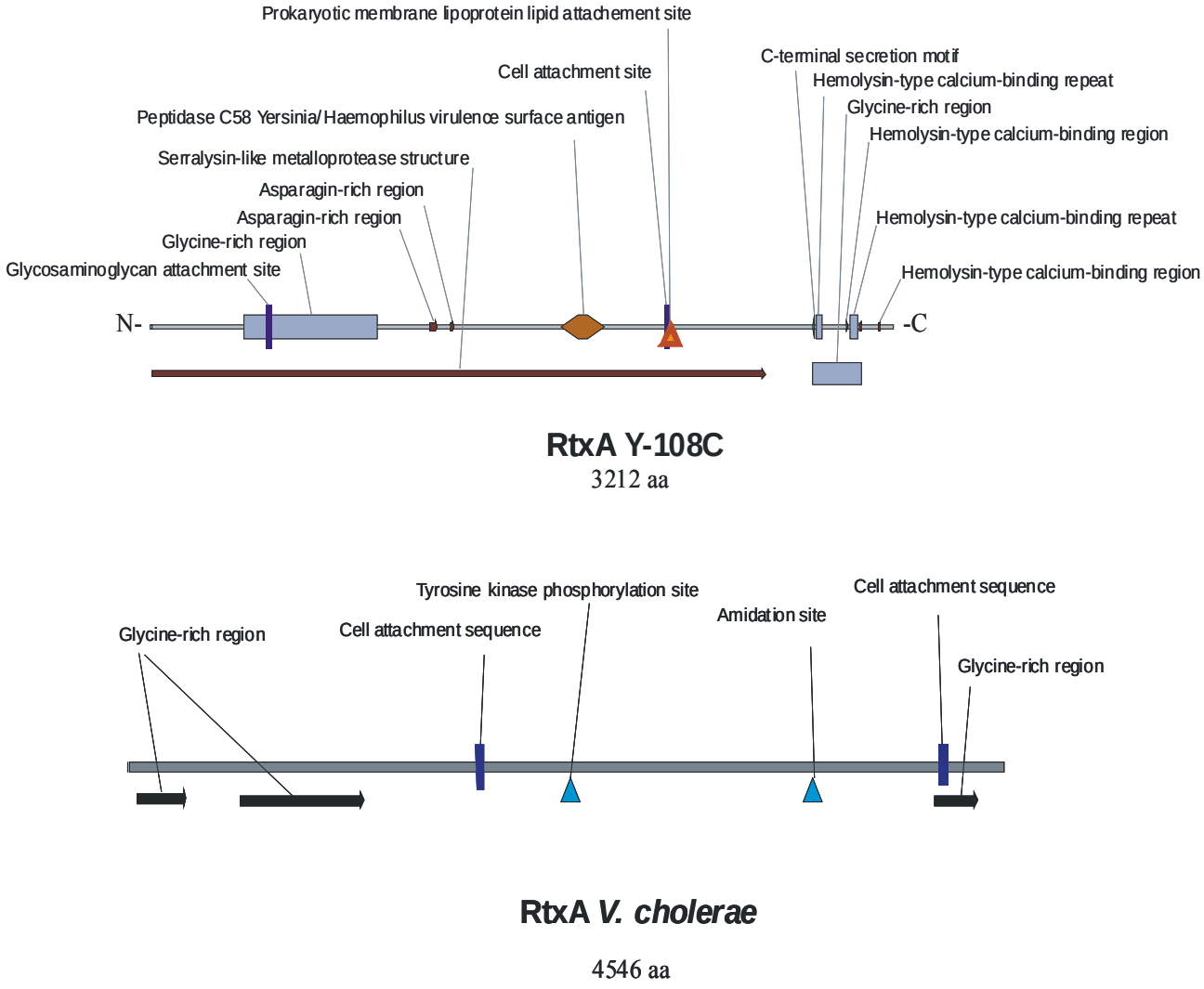


Fig. 14: Schematic representation of the RtxA protein and its features.
Upper figure: map of the RtxA-like toxin from *Y. enterocolitica* Y-108C
Lower figure: map of the RtxA-like toxin from *V. cholerae* El Tor
All putative domains and sites are marked.

The predicted size of RtxA in *Y. enterocolitica* Y-108C is 3,212 aa in size, compared with about 1,000 aa residues for the *E. coli* HlyA or 4,546 aa for the *V. cholerae* RtxA. A pore-forming domain is notably absent in the N terminus of RtxA. In fact, glycine-rich regions of RtxA shows the greatest sequence similarity to other RTX toxins. Normally, they contain hemolysin-type calcium-binding sites which bind calcium. It has been suggested that such internally repeated domain of haemolysins may be involved in Ca-mediated binding to erythrocytes. It has been shown that such a domain is involved in the binding of calcium ions in a parallel beta roll structure (Baumann *et al*., 1993). Such metal-binding sites were found in a group of bacterial exported proteins that includes haemolysin, cyclolysin, leukotoxin and metallopeptidases belonging to MEROPS peptidase family M10 (clan MA(M)), subfamily 10B (serralysin) (Park and Ming, 2002).

Within the enormous length of RtxA, about a dozen sequences conform to the consensus for toxin activation by acylation (Pellett and Welch, 1996). However, whether or how many of these sites are substrates for acylation is unknown. Finally, the C terminus of the protein contains a secretion signal. Although the secretion signals of RTX toxins do not follow a strong consensus, disruption of an "aspartate box" motif, rich in aspartic acid and serine, at the C terminus of other RTX toxins can severely affect secretion (Kenny *et al*., 1992). One RGD (cell attachment) sequence appear at amino acids 2234-2236 of RtxA (Fig. 14). It is interesting, RtxA of *V. cholerae* contains two similar motifs, whereas it is absent in HlyA of *E. coli* (Lin *et al*., 1999). This short peptide sequence, found in fibronectin and other adhesive proteins, facilitates binding to the integrin family of cell surface receptors (Ruoslahti, 1996). What has been called the 'RGD' tripeptide is also found in the sequences of a number of other proteins, where it has been shown to play a role in cell adhesion. These proteins are: some forms of collagens, fibrinogen, vitronectin, von Willebrand factor (VWF), snake disintegrins, and slime mold discoidins. The 'RGD' tripeptide is also found in other proteins where it may also, but not always, serve the same purpose. Whether RGD sequences have any role in activity of RtxA is unknown.

Although large regions of the protein do not share sequence similarity to proteins in the database, certain motifs are present. Such motifs like a glycosaminoglycan attachment site, serralysin-like metalloprotease structure, peptidase C58 *Yersinia*/*Haemophilus* virulence surface antigen domain and prokaryotic membrane lipoprotein lipid attachement site have not been found in other RTX-like toxins (Fig. 14). All of these structures are present in proteins that play a role in a pathogenic process. For example, serralysin is a bacterial Zn-endopeptidase that acts as a virulence factor to cause tissue damage and anaphylactic response

(Park and Ming, 2002). These peptidase include the astacin family, snake venom Zn-endopeptidases, the extracellular metalloproteases from *Serratia* sp., *Pseudomonas* sp. and *Erwinia* sp., and the matrixins. Other protein, *Yersinia*/*Haemophilus* virulence surface antigen, belongs to a group of cysteine peptidases correspond to MEROPS peptidase family C58 (clan CA). They are found in bacteria that include plant pathogens (*Pseudomonas syringae*), root nodule bacteria, and intracellular pathogens (e.g. *Yersinia pestis*, *Haemophilus ducreyi*, *Pasteurella multocida*, *Chlamydia trachomatis*) of animal hosts. Sequences can be extremely divergent outside of a few well-conserved motifs. Members of the family from pathogenic bacteria are likely to be pathogenesis factors.

3.5 Production of recombinant RtxA and generating of a rabbit serum against RtxA

pRTX2100 is a plasmid that contains a 2.1 kb fragment of the *rtxA* gene in pET100/D-TOPO (Invitrogen) vector. The 79 kDa His-tagged product, designed RtxA79, was purified from E. coli BL21 cells in the presence of 6 M Urea on a nickel column (Fig. 15).

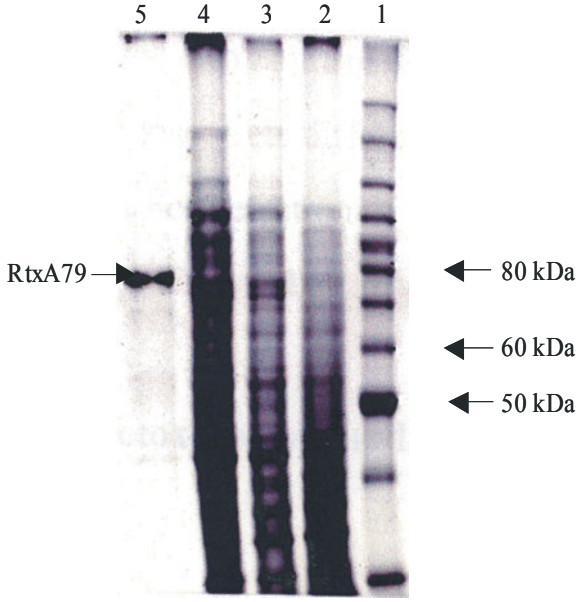


Fig. 15: Purification of the RtxA79. SDS-PAGE (7.5% polyacrylamide)

1. BenchMark Protein Ladder (Invitrogen)
2. Crude bacterial lysate after induction
3. Soluble fraction after disrupting the bacterial pellet through French press
4. Insoluble fraction after disrupting the bacterial pellet through French press
5. RtxA79 after purification on a nickel column

Purified protein was used for immunization of a rabbit and production of the RtxA-specific antibodies.

Sterile culture supernatant fluids and cell lysate from stationary phase cultures of Y11 were examined for the presence of the RtxA by immunoprecipitation and western blotting. RtxA

was immunoprecipitated and subsequently detected using antibody raised against a 79 kDa subfragment of RtxA that had been purified from *E. coli* as a 6xHis-tagged protein. Two weakly visible bands were detected by the RtxA-specific antibody that were not detected by the pre-immune serum (Fig. 16).

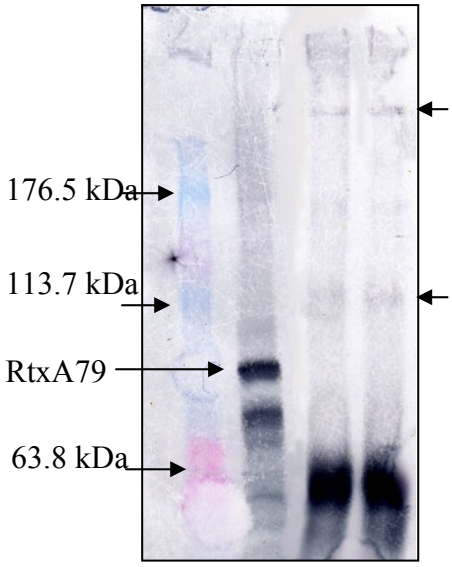


Fig. 16: RtxA immunoblotting:
Antiserum was raised against RtxA79.
Arrows on the right denote the location of RtxA-specific bands.
M – BenchMark Prestained Protein Ladder (Invitrogen)
1- Crude cell lysate of *E. coli* BL21 expressed RtxA79
2- Culture supernatant fluids from Y-108C after immunoprecipitation
3- Crude cell lysate of Y-108C after immunoprecipitation

Together these bands consistent with the predicted full-length size of RtxA.

These data demonstrate that the RtxA protein is synthesized and exported to culture supernatants but in a small quantity. In future, RtxA-specific antibodies may be used for detection antibodies in human serum against RtxA and, if it will be high specific for *Y. enterocolitica* ST O:3, as a tool for serological diagnostic.

4. Structural organization of the pFra virulence-associated plasmid of Rha-positive *Yersinia pestis*

Plague is an acute systemic zoonotic disease caused by infection with *Yersinia pestis.* Evolutionary *Y. pestis* is a recently evolved clone of *Y. pseudotuberculosis* (pathotype) that causes chronic and localized lymphadenitis and gastroenteritis. A group of Rha-positive *Y.*

pestis strains shares certain features of both *Y. pseudotuberculosis* and *Y. pestis*. One of these Rha-positive strains, *Y. pestis* G8786 biovar Antiqua, isolated from a vole in high mountainous Caucasus, Georgia, was compared by suppression subtractive hybridization (SSH) with another Antiqua isolate, *Y. pestis* Yokohama, to uncover lossess and acquisitions in the G8786 genome. Changes in both, chromosomal and plasmid DNA, were detected in G8786.

Several G8786-specific sequences show similarity to genes responsible for the transmissivity of R100 and F plasmids, namely, *traG* (pilus assembly), *traH* (pilus assembly) and *traN* (mating pair stabilization). We supposed that a transmissible replicon might have been acquired by the *Y. pestis* G8786. To prove this, the pFra plasmid and chromosomal DNAs were hybridized with a *traG* probe designed from the sequence of the subtracted fragment. Hybridization studies proved that the *tra*-operon is a part of the pFra plasmid, namely pG8786. To map out the location of the insert of the *tra*-genes in the pG8786 plasmid we have sequenced the complete replicon pG8786.

4.1 General description

Y. pestis strain G8786 was cured of pYV8786 by plating on LB-EGTA agar at 37°C and selecting for the loss of the Cad-phenotype. Loss of the pYV8786 plasmid was proven by plasmid screening and by PCR for the pYV-encoded marker *yopP*. A shotgun library was prepared from pG8786 isolated from a monoplasmid *Y. pestis* G8786 derivative. The entire sequence of pG8786 was determined to be 137,036 bp. Screening and annotation of the sequence with the Pedant-Pro Sequence Analysis Suite (Biomax Informatics AG, Germany) revealed 148 putative coding regions along the entire length of the plasmid. In general, the pG8786 is a pFra plasmid which has additionally acquired *tra* genes (necessary for a conjugational transfer) from an unknown origin but with significant similarity to known and well characterized conjugative plasmids including F and R plasmids (Table 13).

Table 13. List of the characteristics and the closest relatives of the predicted product of each CDS from regions 1 and 2.

CDS/Gene	Position (bp)	Homologue found by BLAST (ORF/Gene, product, organism/element)	Similarity (aa, %)	Accession no.
		region 1 37.546-42.172 bp		
CDS38	38.438-37.648	HCM2.0120c, hypothetical protein, *Salmonella enterica* subsp. *enterica* serovar Typhi, plasmid pHCM2	261/262 (99%)	NP_569592
CDS39	39.737-38.595	HCM2.0121c, putative ribonucleoside-diphosphate reductase beta subunit, *Salmonella enterica* subsp. *enterica* serovar Typhi, plasmid pHCM2	379/380 (99%)	NP_569593
CDS40	42.160-39.845	HCM2.0122c, putative ribonucleoside-diphosphate reductase alpha subunit, *Salmonella enterica* subsp. *enterica* serovar Typhi, plasmid pHCM2	769/771 (97%)	NP_569594
		region 2 81.956-114.573 bp		
traA	82.124-82.480	*traA*, prepropilin, *Escherichia coli* K-12 strain CR63, plasmid F	63/69 (91%)	NP_061453
traL	82.631-82.936	*traL*, TraL, function - F pilus assembly, *Escherichia coli* K-12 strain CR63, plasmid F	87/99 (87%)	NP_061454
traE	82.951-83.517	*traE*, TraE, function - F pilus assembly, *Escherichia coli* K-12 strain CR63, plasmid F	159/188 (84%)	NP_061455
traK	83.504-84.247	*traK*, TraK, function - F pilus assembly, *Escherichia coli* K-12 strain CR63, plasmid F	173/224 (77%)	NP_061456
traB	84.225-85.628	*traB*, TraB, function - pilus assembly, *Escherichia coli* K-12, plasmid R100-1	362/467 (77%)	AAB07770
traV	85.650-86.189	*traV*, TraV, function - pilus biogenesis, *Shigella flexneri* 2b strain 222, plasmid R100	114/163 (69%)	NP_052954
traV	86.267-86.533	*traR*, TraR, conjugative transfer protein, *Salmonella typhi*, plasmid pED208	55/88 (62%)	AAM90708

Table 13. Continued

CDS/Gene	Position (bp)	Homologue found by BLAST (ORF/Gene, product, organism/element)	Similarity (aa, %)	Accession no.
orfX1(CDS81)	86.537-86.722	*orfX1*, hypothetical protein, transfer region, *Salmonella typhi*, plasmid pED208	54/60 (90%)	AAM90709
orfX2	86.604-87.110	*orfX2*, hypothetical protein, transfer region, *Salmonella typhi*, plasmid pED208	73/95 (76%)	AAM90710
traP	87.100-87.666	*traP*, conjugative transfer protein, *Salmonella typhimurium* LT2, plasmid pSLT	62/137 (45%)	NP_490568
traC	87.659-90.290	possible pseudogene *traC*, *Escherichia coli* K-12 strain CR63, plasmid F	301/339 (88%)	NP_061463
trbI	90.287-90.634	*trbI*, pilus assembly protein, *Salmonella typhimurium* LT2, plasmid pSLT	78/91 (85%)	NP_490574
traW	90.631-91.281	*traW*, pilus assembly protein, *Salmonella typhimurium* LT2, plasmid pSLT	143/187 (76%)	NP_490575
traU	91.278-92.270	*traU*, pilus assembly protein, *Salmonella typhimurium* LT2, plasmid pSLT	276/320 (86%)	NP_490576
trbC	92.462-92.917	*trbC*, hypothetical protein, *Shigella flexneri* 2b strain 222, plasmid R100	111/151 (73%)	NP_052965
traN	92.914-94.764	*traN*, mating pair stabilization protein, *Salmonella typhimurium* LT2, plasmid pSLT	455/607 (74%)	NP_490579
traF	95.017-95.766	*traF*, TraF, *Escherichia coli*, plasmid R100-1	193/234 (82%)	AAB61943
traO	95.693-96.031	*traQ*, pilin chaperone, *Salmonella typhimurium* LT2, plasmid pSLT	47/63 (74%)	NP_490582
trbB	95.970-96.566	*trbB*, hypothetical protein, *Escherichia coli* K-12 strain CR63, plasmid F	139/179 (77%)	NP_061474
traH	96.563-97.936	*traH*, pilus assembly protein, *Salmonella typhimurium* LT2, plasmid pSLT	394/446 (88%)	NP_490584
traG	97.936-100.755	*traG*, responsible for pilus biogenesis and stabilization of mating pairs, *Shigella flexneri* 2b strain 222, plasmid R100	761/939 (81%)	NP_052976

Table 13. Continued

CDS/Gene	Position (bp)	Homologue found by BLAST (ORF/Gene, product, organism/element)	Similarity (aa, %)	Accession no.
Y1062	102.083-100.876	Y1062, possible pseudogene, putative transposase IS*285*, *Yersinia pestis* KIM 10+, plasmid pMT-1	381/381 (100%)	AAC82722
traD	102.087-104.276	*traD*, TraD, *Escherichia coli* F sex factor	547/705 (77%)	AAC44181
traI	104.276-109.516	*traI*, TraI, function - oriT nicking and unwinding, *Escherichia coli* K-12 strain CR63, plasmid F	1324/1759 (75%)	NP_061483
traX	109.380-110.255	*traX*, F pilin acetylation protein, *Shigella flexneri*, plasmid pWR501	160/284 (56%)	NP_085416
finO	110.322-111.041	*finO*, fertility inhibition protein, *Escherichia coli*, plasmid F	75/180 (41%)	P22707
bcfH	111.028-111.825	*bcfH*, putative thiol-disulfide isomerase, *Salmonella typhimurium* LT2	138/244 (56%)	NP_459033
nuc	111.973-112.527	*nuc*, endonuclease, *Yersinia enterocolitica* strain 15673 (biogroup1A/serogroup 0:5) pYV plasmid	135/174 (77%)	CAA73744
copB	112.418-112.648	*copB*, CopB, function - replication control, *Enterobacter intermedius*, plasmid pLV1402	36/46 (78%)	CAA08928
copA	112.835-112.750	*copA*, small RNA, function – antisense control, *Escherichia coli*, plasmid R1	71/81 (87%) (bp)	V00326
tapA	112.867-112.944	*tapA*, TapA, function - antisense control, *Enterobacter intermedius*, plasmid pLV1402	24/25 (96%)	CAA08929
repA	112.925-113.800	*repA*, replication protein, *Enterobacter intermedius*, plasmid pLV1402	234/287 (81%)	CAA08930
SMR0139	114.351-114.847	SMR0139, possible pseudogene, ATP/GTP-binding protein, *Serratia marcescens*, plasmid R478	143/164 (87%)	NP_941209

When the pG8786 ORFs had similarity to known proteins in the database, we assigned the putative protein a likely function. A total of 62 of these ORFs are transcribed in a clockwise orientation, while the remaining 86 ORFs are transcribed counterclockwise. All putative ORFs have significant homology to previously described hypothetical or characterized

proteins in the GenBank database. 79% of them exactly matched ORFs of plasmid pMT1 of *Y. pestis* biovar Mediaevalis strain 91001 (Accession no. NC_005815). The position and transcriptional orientations of all ORFs are shown in Figure 17. In contrast to other sequenced pFra replicons (pMT1 from *Y. pestis* biovar Mediaevalis strains KIM5 and KIM10+, and pMT1 from *Y. pestis* biovar Orientalis strain CO92) this replicon contains two additional large coding regions: i) three ORFs with high similarity to the HCM2.0120c, HCM2.0121c, and HCM2.0122c genes of plasmid pHCM2 from *Salmonella enterica* serovar Typhi strain CT18 (denoted region 1 in Figure 19) and ii) a large cluster of transfer genes (region 2 in Figure 19). We have not found any crucial deletions in the pFra part of pG8786 besides the absence of two copies of the IS*100* element present in plasmid pMT1 of *Y. pestis* biovar Mediaevalis strain 91001 (Table 14).

Two potential regions of plasmid replication and one partitioning system were discovered on pG8786. One region of replication originates from the pFra plasmid (Hu *et al*., 1998; Lindler *et al*., 1998), while the second one has high similarity to the alpha replicon pLV1402 plasmid of *Enterobacter intermedius* (Osborn *et al*., 2000). The plasmid partitioning function was identical to the *parABS* system of pFra (Hu *et al*., 1998; Lindler *et al*., 1998).

The overall G+C content of pG8786 was 51.96% in comparison to the lower overall G+C content (47.64%) of the chromosome of *Y. pestis* KIM or CO92 (Deng *et al*., 2002; Parkhill *et al*., 2001) or pMT1 (50.2%) (Hu *et al*., 1998). Surprisingly, region 2 covering nucleotides 81,956 to 114,573 has a G+C content of 57.58% (Fig. 18) that is much higher than the overall G+C content of the backbone of the plasmid pointing to its horizontal acquisition.

pG8786 contains two copies of the IS*200*-like element (also known as IS*1541* in *Y. pestis*) in opposite orientation. The first IS*200* insertion is located between positions 34,588 and 35,076 and the second one is close to the first one (positions 36,944 to 37,453). Also two copies of the IS*285*-like element were found in opposite orientation (positions 80,712 to 81,920 and 102,083 to 100,876, respectively).

The second IS*285* had a frameshift after codon 169 and thus appeared to be a non-functional remnant. One copy of each, an IS*1328*-like and an IS*100* element, was found on pG8686 (Table 14 and Fig. 18). They have the same orientation as the first copy of IS*285*.

Table 14. Distribution of IS-elements in five sequenced pFra plasmids of *Y. pestis*

Y. pestis strain	Number of IS*100* copies	Number of IS*285* copies	Number of IS*1541* (IS*200*) copies	Number of IS*1328*-like copies	Accession no.
91001	3	1	1	1	NC_005815
KIM10+	2	1	1	1	AF074611
KIM5	2	1	1	1 (IS*1618*)	AF053947
CO92	2	1	1	0	NC_003134
G8786	1	2	2	1	AJ698720

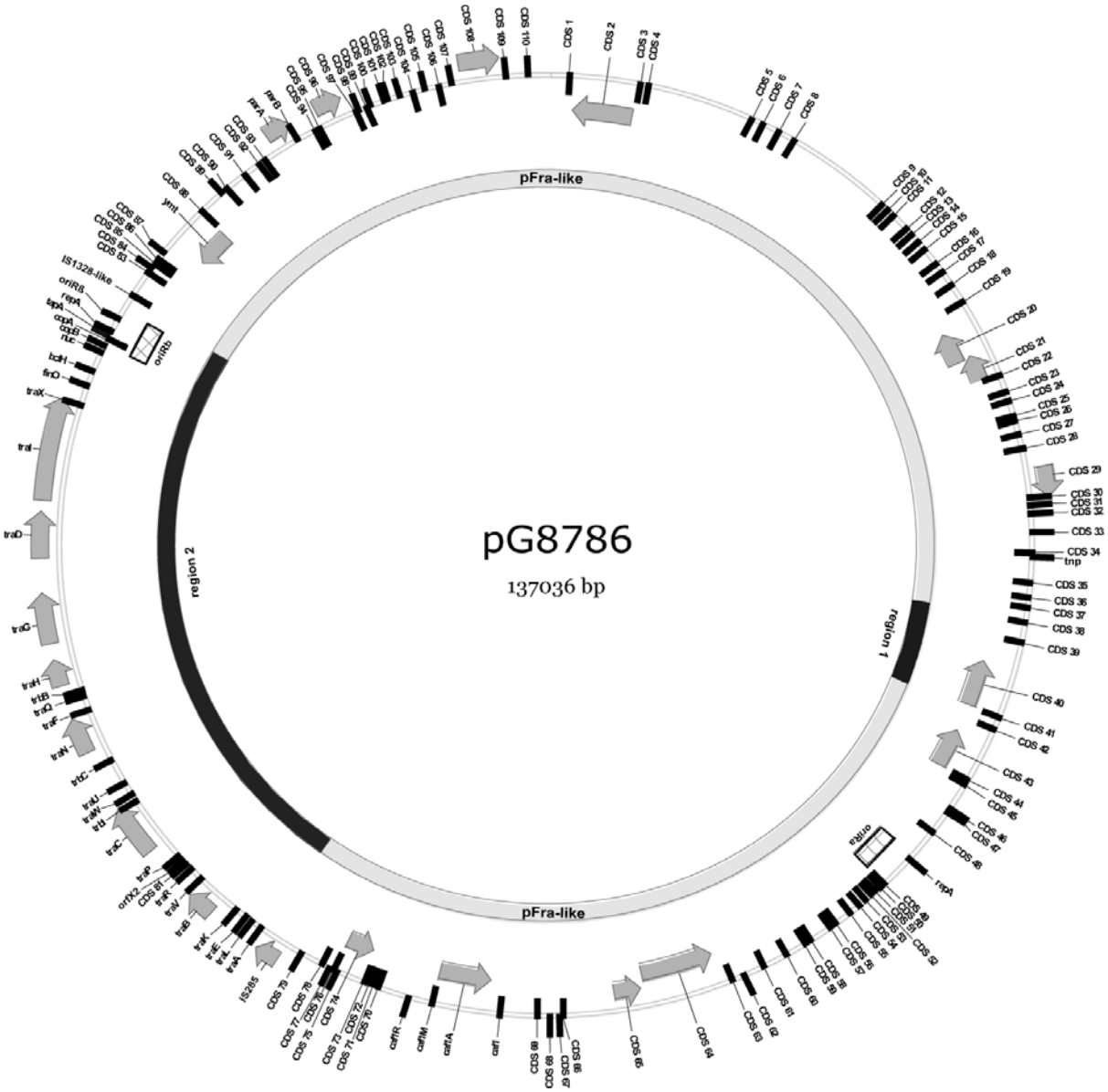


Fig. 17: Map of the pG8786 plasmid. The internal circle depicts region 1, region 2 and the pFra-like backbone. The outer circle shows ORFs and their orientation which are denoted by their positions: arrows and boxes outside of the ring indicate clockwise transcription, and inside the ring indicate counterclockwise. The map was derived from the annotated DNA sequence by the Vector NTI (InforMax, USA) computer program and edited in CorelDRAW.

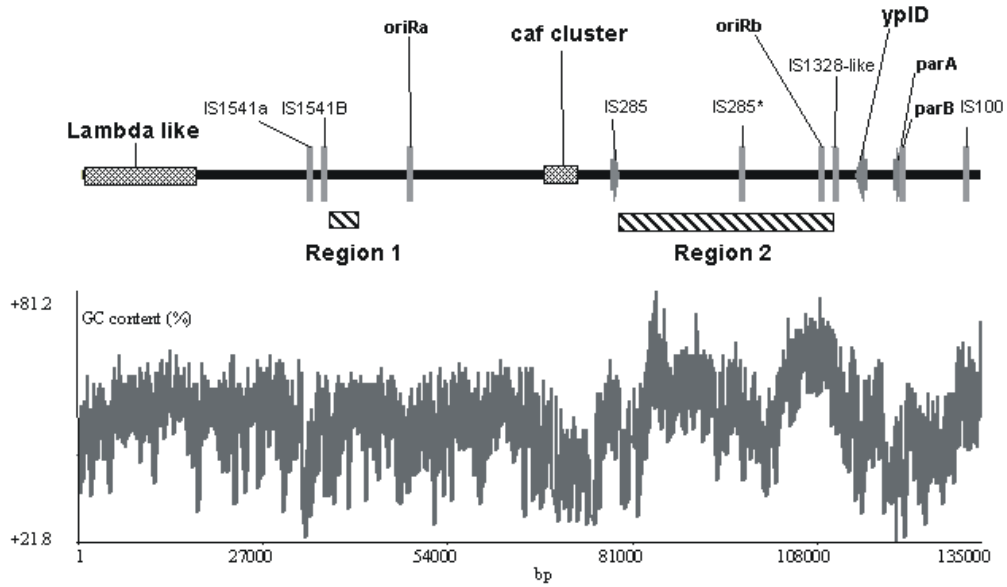


Fig. 18: The G+C content and the graphic map of pG8786. The plot showing the G+C content was derived by the Vector NTI (InforMax, U.S.A.) program. The plot displays selected ORFs and some other annotated features to the correct scale (the upper figure). The scale below the GC plot displays the size of the plasmid in base pairs. IS285* - IS*285*-insertion sequence which appeared to be a non-functional remnant.

4.2 ORFs of region 1

The 4,626-bp region 1 is absent from the pFra plasmids of *Y. pestis* KIM and CO92 representing biovars Mediaevalis and Orientalis, respectively (Fig. 19). Surprisingly this region is present in the pFra plasmid from the avirulent *Y. pestis* biovar Mediaevalis strain 91001 isolated from *Microtus brandti* in Inner Mongolia, China (Accession no. AE017045). It is also 96% identical to the plasmid pHCM2 of *Salmonella enterica* serovar Typhi strain CT18 (Parkhill *et al*., 2001). Our analysis revealed three putative ORFs (CDS38, CDS39, and CDS40) spanning bp 37,641 to 42,160 (Table 13). CDS38 is highly similar to the HCM2.0120c hypothetical protein. CDS39 and CDS40 are putative beta and alpha ribonucleoside-diphosphate reductase subunits which might be necessary for the deoxyribonucleotide metabolism.

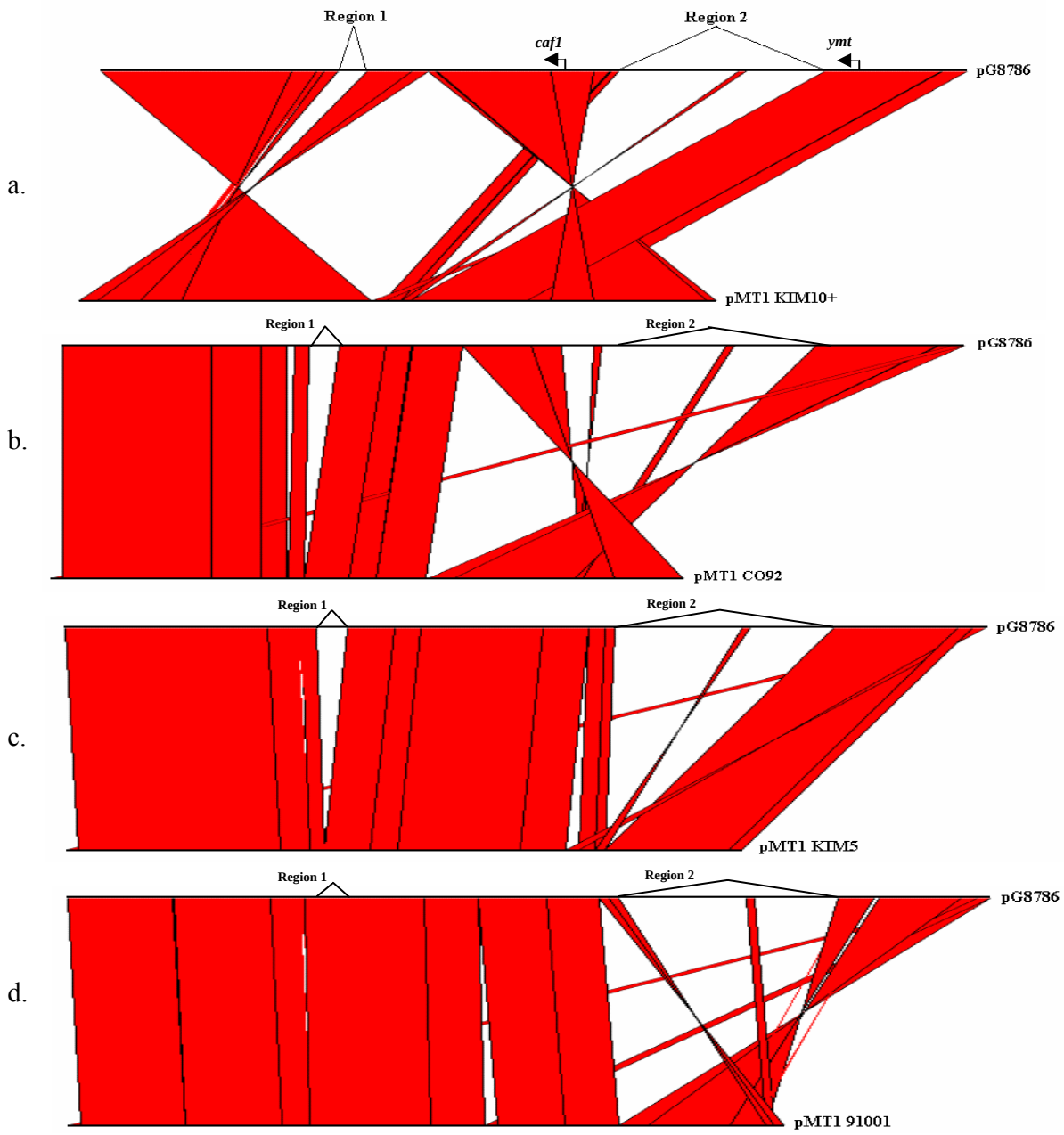


Fig. 19: Graphical comparison of different pFra plasmids with pG8786 derived by the ACT (Artemis Comparison Tool) program (The Wellcome Trust Sanger Institute, UK). a) pG8786 vs. pMT-1 of *Y. pestis* KIM10+; b) pG8786 vs. pMT1 of *Y. pestis* CO92; c) pG8786 vs. pMT1 of *Y. pestis* KIM5; d) pG8786 vs. pMT1 of *Y. pestis* 91001. Areas of pG8786 that are not present in the other pFra plasmids are shown as the region 1 and region 2. Vertical lines show similar parts of the plasmids.

4.3 ORFs of the transfer region

ORFs spanning bp 81,956 to 111,038 (region 2) in pG8786 were found to be similar to the *tra* region genes of the F-like plasmids which belong to the type IV family of secretion systems (T4SS; Table 13). Analysis of the pG8786 *tra* region further showed 25 putative ORFs (*traA-bcfH*), whereas the F *tra* region has 37 ORFs (Frost *et al*., 1994). However, the cryptic conjugative plasmid from *Y. enterocolitica* 29930 contains even less genes i.e. 16 ORFs (Strauch *et al*., 2003). The *tra* region of pG8786 is similarly organized and highly homologous to the *tra* regions of F-like plasmids from different incompatibility groups of the IncF family – IncFI (F), IncFII (R-100, R100-1) and IncFV (pED208) (Table 13, Fig. 20a). Accordingly, the putative pG8786 transfer genes were designated by their homologs in the IncF plasmids.

The gene products can be organized into four groups based on functions inferred from their closely related homologs: pilus biogenesis (TraA-V, TraW, -U, TrbC, TrbI, TraF, -H, -Q, -X and the N-terminal region of TraG); regulation (FinO); DNA nicking and initiation of transfer (TraI, TraD), and mating-aggregate stabilization (TraN, -G). Other non-classified components of the *tra* region are TraP, a protein that stabilizes the extended pilus, TrbB, a putative thioredoxin homolog, and hypothetical proteins OrfX1 and OrfX2 (Frost *et al*., 1994; Lawley *et al*., 2003; Lu *et al*., 2002).

Surprisingly, the genes with similarity to *traM* (the function of signaling that DNA transfer should begin), *traJ* (a positive regulator of transcription of the *tra* operon), *traT*, and *traS* (surface exclusion) were not detected in the *tra* region of pG8786. We have also identified only the 3'-end remnant of *traY* located next to the first copy of IS*285*. For the reason that the gene organization of the *tra* region of pG8786 mostly resembles the transfer region of plasmid pED208 of *Salmonella typhi* (Fig. 20a, 21) (Lu *et al*., 2002), we speculate that genes *traM*, *traJ*, *traY*, *traT* and *traS* might be deleted or truncated in the case of *traY*.

The *oriT* region is arbitrarily defined as the region at the beginning of the *traM* gene. This region contains the site where nicking occurs and transfer of the single-stranded DNA is initiated, in a 5'-to-3' manner, into the recipient cell (Frost *et al*., 1994). As the *traM-Y* gene locus was absent from pG8786, we tried to define a possible origin of transfer (*oriT*). However, an expanded search of the pG8786 sequence did not reveal any region of the plasmid that might function as *oriT*. Nevertheless, plasmids without defined *oriT* have been described (Galli *et al*., 2001).

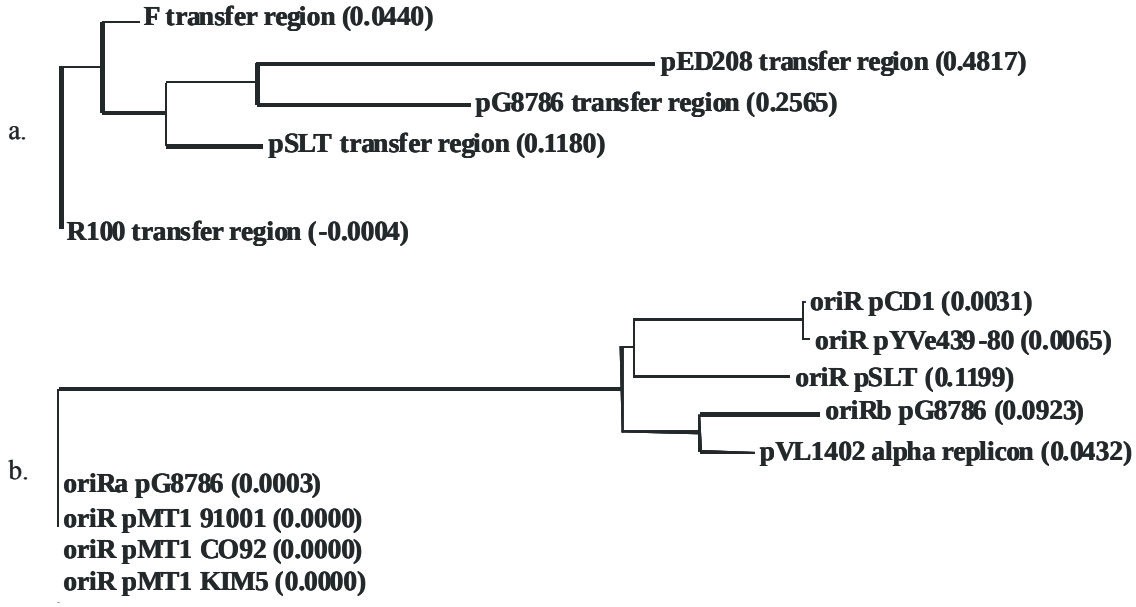


Fig. 20: a) The phylogenetic tree of the transfer regions from various origins. b) The phylogenetic tree of the replication origins from different microorganisms. The trees are built using the Neighbor Joining method (NJ) of Saitou and Nei by the Vector NTI program (InforMax, U.S.A.). AlignX displays the calculated distance values in parenthesis following the molecule name displayed on the tree.

An interesting feature of the pG8786 *tra* region is the presence of the gene corresponding to *finO* (Table 13). FinO is a part of the FinOP system that is a key determinant defining the frequency of IncF plasmid-mediated DNA transfer. TraJ, a positive regulator of the *tra* genes is controlled at the post-transcriptional level by two negative elements, *finP* and *finO*. FinP is a plasmid-specific antisense RNA, whereas *finO* encodes a co-repressor, which is almost identical (over 95% identity) and cross-reactive among various F-like plasmids (Frost *et al*., 1994). We suppose that the intact FinO of pG8786 can also repress the transcription of *traJ* of other F-like conjugative plasmids which might have been acquired by G8786 cells. Thus the frequency of transfer of such an acquired plasmid might be dramatically reduced.

To check self-transmissivity of pG8786, we have inserted a chloramphenicol gene cassette between the *caf1* gene and CDS69 (primers' position bp 69,955 – bp 70,412) to tag this plasmid. *Y. pestis* G8786 (pG8786-Cm^R) was mated with recipient *E. coli* JM109 (Nal^R). However, we did not detect a transfer of the Cm^R marker and thus of the labelled plasmid as well. Also our attempts to mobilize pG8786-Cm^R with the broad range RP4 IncP-alpha plasmid (Pansegrau *et al*., 1994) were unsuccessful. Nevertheless, it cannot be excluded that

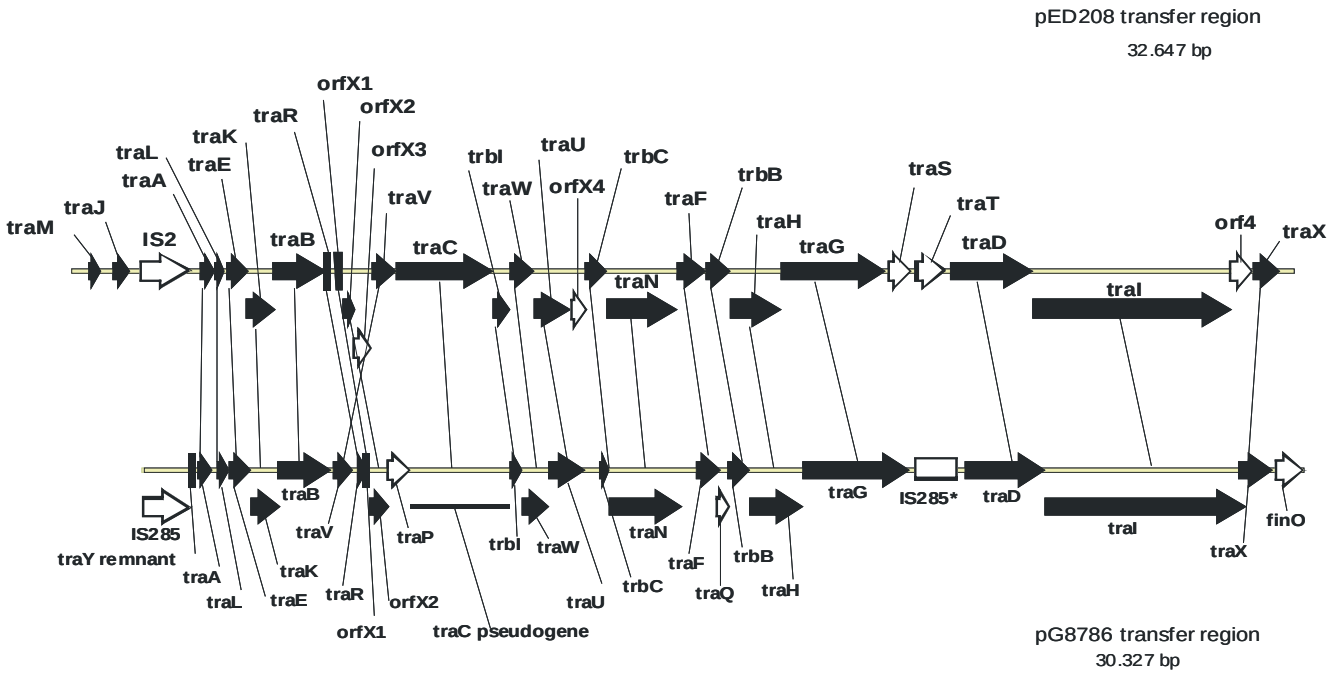


Fig. 21: Line-up of the *tra* genes of pG8786 (lower figure) with the *tra* genes of the pED208 from *Salmonella enterica* serovar Typhi (upper figure).
The black arrows point out ORFs which are equally represented in both plasmids. The white arrows represent ORFs which are absent in either of the two transfer regions. IS285* - IS*285*-insertion sequence which appeared to be a non-functional remnant. The vertical lines between arrows show similar ORFs.

pG8786 might be transmissible if supplemented with the missing *tra* genes in trans or if mated with a more suitable recipient strain.

4.4 Replication and plasmid maintenance

As mentioned above, DNA sequence analysis revealed two potential origins of plasmid replication which were named *oriRa* (bp 49,713 to 51,622) and *oriRb* (bp 112,331 to 114,155) (Fig. 17, 18). *oriRa* is identical with the replication origin of pFra which is also similar to RepFIB, RepHI1B, and P1 and P7 replicons (Lawley *et al*., 2003).

The second replication origin *oriRb*, which was localized in region 2, showed very high similarity (89% in bp) to the alpha replicon (RepFIIA) of pLV1402 of *Enterobacter intermedius* (Osborn *et al*., 2000). It is closely related to the IncFII virulence-associated replicons of pCD1 of *Y. pestis* (AF074612) and pYVe439-80 of *Y. enterocolitica* (M55182) (Fig. 20b). On the basis of the similarity to these replicons, the following genetic features were identified: *copB* (bp 112,331 to 112,648), *copA* (bp 112,835 to 112,750), *tapA* (bp 112,867 to 112,944), *repA* (bp 112,925 to 113,800) and *oriRβ* (bp 113,972 to 114,155) (Table 13, Fig. 17). In the IncFII replicons an antisense RNA molecule (CopA) inhibits synthesis of the replication protein (RepA) by binding to the leader region of the *repA* mRNA (CopT). RepA synthesis depends on translation of a short leader peptide (TapA) that is not expressed when CopA binds to CopT, thereby preventing translation of RepA and consequently preventing replication of the plasmid (Blomberg *et al*., 1992). The sequence 5'-TTGCCCACA-3', which may function as a binding site for the protein DnaA could be defined 174 bp downstream of *repA*. This sequence matches the DnaA box in seven of nine positions (Praszkier *et al*., 1991).

D. DISCUSSION

1. Method of subtractive hybridization to identify genomic differences among *Yersinia* species

The method of suppressive subtractive hybridization (SSH) has been successfully applied to map out genomic differences between closely related yersiniae, for example the work of Iwobi (Iwobi *et al*., 2002) who used SSH to uncover genomic differences between highly and nonpathogenic serotypes of *Y. enterocolitica*. In another study, comparison of *Y. pestis* and *Y. pseudotuberculosis* revealed seven DNA regions in *Y. pestis* that do not occur in *Y. pseudotuberculosis*, with four of them mapping to the same region on the *Y. pestis* genome (Radnedge *et al*., 2002). Subtractive hybridization was also successfully applied for whole genome comparisons between different strains of *Y. pestis* with the goal to develop signatures for epidemiological studies (Radnedge *et al*., 2001). Six species-specific difference regions (DFRs) were identified between different biovars (Antiqua, Mediaevalis and Orientalis) of *Y. pestis*. The DFRs were mapped and four were flanked by insertion sequences. The appearance of these DFRs in eighty geographically diverse strains of *Y. pestis* representing all three biovars was determined and revealed genomic plasticity resulting from the acquisition and deletion of these DNA regions. Additionally *Y. pestis* biovar Orientalis was found to possess DFR profiles different from Antiqua and Mediaevalis biovars, reflecting most probably the recent origins of this biovar.

This work represents the attempt to find novel, potential virulence-associated sequences in weakly pathogenic *Y. enterocolitica* strains as well as to identify genomic features of Rha-positive *Y. pestis* strains using subtractive hybridization.

1.1. SSH applied for *Y. enterocolitica* starins

To map out genomic differences between highly pathogenic *Yersinia enterocolitica* WA-C BG 1B, ST O:8 strain and weakly pathogenic *Y. enterocolitica* Y-108C BG 4, ST O:3 strain we have applied a method of suppression subtractive hybridization (SSH). In total, 428 WA-C-specific and 83 Y-108C-specific sequences were uncovered by SSH. About 53% of O:8 and 49% of O:3 subtracted fragments were tester-specific. Half of the tester-specific fragments were highly homologous (70-95% identity) to the sequences of the *Yersinia pestis* CO92 genome. These sequences represented known genes from several groups: (1) genes

involved in O-antigen biosynthesis, (2) host-specific restriction-modification systems, (3) systems of iron and heme acquisition and storage, (4) flagellar biogenesis genes, (5) putative virulence factors, (6) drug resistance genes, and (7) mobile elements. The large fraction of sequences homologous to mobile genetic elements may reflect their contribution to dispersing putative virulence traits and to the ongoing rearrangements of genetic islands. The spectrum of known virulence genes and novel virulence-associated DNA fragments was rather bioserotype-specific, than strain specific. Of note, part of the tester-specific sequences (in total 45 sequences) displayed either low homology (65%) or no homology to the known sequences. Unexpectedly, we have found several putative virulence determinant, for example *rtxA*-like putative cytotoxin, *xnp2*, nematocidal toxin and components of a new type III secretion system in the *Y .enterocolitica* O:3 genome. We speculate that these pathogenic factors might be a cause of different virulence level in weakly (BG 2-5) and nonpathogenic (BG 1A) biogroups of Y. *enterocolitica*. In summary, this study provides further evidence for the considerable diversity of the genomes within the *Y. enterocolitica* species. We detected more than 500 novel DNA fragments of *Y. enterocolitica* genomes, part of which are highly virulence-associated and may represent new targets for diagnostic purposes or epidemiological study.

2. IS*1331* is a novel insertion sequence element which is specific to the weakly pathogenic European biogroups and serotypes of *Y. enterocolitica*

2.1 IS*1331* belongs to the IS*21* family

We have reported on a first insertion sequence element, IS*1331*, characterized in the weakly pathogenic *Y. enterocolitica* ST O:3, ST O:1 and ST O:2 strains. Based on the differences between the nucleotide content of IS*1331* copies we have revealed two IS*1331* isoforms: IS*1331A* and IS*1331B*. Comparison of IS*1331* genes, *istA* and *istB*, with GenBank database indicated that IS*1331* belongs to the IS*21* family. Members of this family have the length range 1,950-2,500 bp and are therefore among the largest bacterial IS elements. They exhibit two consecutive ORFs: a long frame designated *istA* (transposase) and a shorter downstream frame, *istB* (NTP-binding protein). Usually members of the IS*21* family have terminal inverted repeats whose lengths may vary between 11 (IS*21*) and 50 bp (IS*5376*) and generally terminate in the dinucleotide 5'-CA-3'. Insertion of these elements results in a direct target duplication of 4 or, more frequently, 5 bp while two members (IS*53* and IS*408*) may generate

8 bp repeats (Mahillon and Chandler, 1998). The IstA of IS*21* family elements usually carries a motif related to the widespread integrase DD-E motif and a potential helix-turn-helix motif. The IstB carries a relatively well conserved potential NTP-binding domain. Overall identities range from 10 to 59% for IstA and from 25 to 67% for IstB (Mahillon and Chandler, 1998; Berger and Haas, 2001). Members of the IS*21* family are widespread. Isoforms of IS*21* (IS*640*, IS*21p*, IS*100*) have been described in *Y. pestis* and *Y. pseudotuberculosis* (Hu *et al*., 1998; Lohe *et al*., 1996; Filippov *et al*., 1995; Podladchikova *et al*., 1994). The presence of the helix-turn-helix motif and the DD-E motif in the putative transposase IstA and the ATP/GTP binding P-loop motif in the putative NTP-binding protein IstB of IS*1331* confirms its inclusion into the IS*21* family. Nevertheless, despite the fact that IS*1331* has several characteristics of the IS*21* family repetitive elements, we have not found any direct repeats in the target sequence after IS*1331* transposition. One of the IS*1331* copies, which has the 5 bp deletion in *istA* gene, might be non-functional due to the frameshift (Fig. 3C, variant 4)

2.2 IS*1331* can promote diverse genomic rearrangements

Selfish IS elements can promote various genomic rearrangements including deletions, inversions as well as insertions of foreign unrelated sequences supplying the host chromosome with movable regions of homology. Recombination between such homologous regions might result in horizontal transfer of foreign genetic information and its establishing in a new host. IS*1331* is present in a single (serotypes O:1, O:2, O:5,27) or multiple (serotypes O:3, O:9) copies on the chromosome of European pathogenic yersiniae (Fig. 7). However, a single IS*1331* copy is localized on the pYV virulence plasmid of BG 2 and BG 4 *Y. enterocolitica* strains. With a second copy of the IS element present on the chromosome a homologous recombination can occur resulting in a transient pYV integration into the bacterial chromosome. Such an event was observed with IS*100*-mediated integration of the pCad and pFra virulence-associated plasmids in *Y. pestis* (Protsenko *et al*., 1991). Also presence of several chromosomal IS*1331* copies might increase the fluidity of *Y. enterocolitica* genome.

2.3 Several copies of IS*1331* are located on an uncharacterized phage

IS*1331* is integrated into different locations on the chromosome of *Y. enterocolitica*. The type of the genes in the neighbourhood of IS*1331* (6 of 10 IS*1331* flanking sequences have similarity to putative phage proteins, Table 9) may speak in favour of a presence of an

uncharacterized phage in *Y. enterocolitica* low pathogenic serotypes. Also it seems probable that IS*1331* might be originally located on such a phage and later has transposed to pYV-like plasmids and chromosome of its new hosts.

2.4 IS*1331* might increase the expression of the downstream genes

Promoter sequences in the IS*1331* inverted repeats might serve as additional movable promoters for the genes suffered IS*1331* insertion. IS*1331* is inserted in ORF181-ORF155 intergenic region of the pYV-like plasmids of *Y. enterocolitica* O:3 (Y11 and Y-108P strains) and JD E029 O:1. ORF181 and ORF155 are putative ORFs encoding hypothetical proteins with the unknown function. It is interesting that ORF181 is absent in two sequenced pYV-like plasmids – pYVa127/90 and pYVe8081 from *Y. enterocolitica* A127/90 and 8081 strains (Acc. No. NC_004564 and NC_005017, respectively) that belong to the highly pathogenic serotype O:8, but is present in pYV-like plasmids from the weakly pathogenic serotypes O:9 (pYVe227, Acc. No. NC_002120), O:3 and O:1 (present study). On the other hand, ORF155 is highly conserved and located in all sequenced pYV-like plasmids just before *yopO*, which encodes a serine kinase (Galyov *et al*., 1993). Also, in the case of *Y. enterocolitica* JD E029 O:1 this intergenic space contains three other ORFs: *yfc*, *yrc* and *tnpA*. Based on these data we speculate that ORF155 might be a part of the Yop virulon and important for the virulence or plasmid maintenance. Also potential promoter sequences in IS*1331* inverted repeats might increase the expression of the downstream genes, for example ORF155.

2.5 IS*1331* is restricted to human and animal weakly pathogenic European *Y. enterocolitica* bioserotypes

Surprisingly, IS*1331* is absent not only in the strains of the American biogroup 1B, but also in the strains of the non-pathogenic European biogroup 1A. It might be an additional evidence for the distant ecological and evolutionary relationships between weakly and non-pathogenic European serotypes.

The novel IS element is restricted to human and animal weakly pathogenic European *Y. enterocolitica* pathotypes, which are proposed to form a new species *Y. enterocolitica palearctica* in contrast to *Y. enterocolitica* 1B strains making up *Y. enterocolitica enterocolitica* (Neubauer *et al*., 1999). The presence of IS*1331* as a signature of European weakly pathogenic bioserotypes of *Y. enterocolitica* suggests its possible application in diagnostics and epidemiology of *Y. enterocolitica* infections.

Although members of the IS*21* group are found in closely related *Y. pestis* and *Y. pseudotuberculosis* they are not present in American *Y. enterocolitica* biotypes, which carry multiple IS elements but of the other IS groups, such as IS*3*, IS*605*, IS*110* and IS*4* (see *Introduction*). Thus distribution of the IS elements of certain groups in two *Y. enterocolitica* pathotypes is in favour of the parallel evolution of American and European *Y. enterocolitica* strains.

3. Identification of a new putative toxin, RtxA, in *Y. enterocolitica* Y-108C

We identified a new putative toxin, RtxA, of *Y. enterocolitica* subsp. *palearctica* that belongs to the family of RTX (repeat in toxin) toxins by using a combination of genomic sequence analysis and representational difference analysis. The *rtxA* gene was uncovered by SSH and subsequently sequenced. The *rtxA* gene was found to be specific for *Y. enterocolitica* subsp. *palearctica*, where it was found in a single copy in the virulent strains Y-108C and in a further clinical isolates. The *rtxA* gene, in contrast, was not found in non-pathogenic European and highly pathogenic American biogroups and serotypes.

3.1 Features of Rtx-like toxins

The Rtx toxins comprise a family of large, heat-labile, Ca^{2+}-dependent, pore-forming cytotoxins secreted by a wide variety of Gram-negative humanand animal pathogens. This family includes hemolysins of *E. coli* (HlyA) (Felmlee *et al*., 1985) and *Actinobacillus pleuropneumoniae* (Chang *et al*., 1989), *Bordetella pertussis* adenilate cyclase-hemolysin (CyaA) (Glaser *et al*., 1989), leukotoxins of *Pasteurella haemolitica* (LktA) (Lo *et al*., 1987), cytotoxin RtxA of *Vibrio cholerae* El Tor (Lin *et al*., 1999), cytotoxin RtxA of *Legionella pneumophila* (Cirillo *et al*., 2000) and Fe-regulated Rtx-like toxin of *Neisseria meningitidis* (Thompson *et al*., 1993). Most Rtx toxins are proteins with a molecular mass of 100-200 kDa and are post-translationally activated by acylation via a specific activator protein. The repeated structure of RTX toxins, which gave them their name, is composed of several glycine-rich nonapeptides on the C-terminal half of the protein that bind Ca^{2+}. The toxic activity of Rtx toxins in host cells may lead to necrosis and apoptosis, but the underlying detailed mechanisms are currently under investigation (Frey and Kuhnert, 2002). The hemolytic activity is the most obvious phenotype of many Rtx toxins under *in vitro* conditions. The glycin-rich repeats have a strong capacity to bind Ca^{2+}, which was shown to be involved in binding to erythrocytes together with the lapidated amino acid residues

(Hughes *et al*., 1992). However, the erythrocytes might not be the main targets of Rtx toxins. *In vivo*, Rtx toxins induce inflammatory mediators or exert cytotoxic and cytolitic effects mainly to cells of the host's immune defense, thus provoking necrosis, apoptosis, inflammation and disease (Czuprynski and Welch, 1995). Rtx toxins are supposed to act in synergy with lipopolysaccharides (Czuprynski and Welch, 1995). The pore-forming activity was shown for several Rtx toxins, but was studied in detail mostly for the *E. coli* α-haemolysin HlyA (Benz *et al*., 1992). The 30-50 most C-terminal amino acids of a structural toxin protein constitute the signal for the specific type I secretion system which is encoded by two genes, both located on the same operon as the activator gene (acyltransferase), and the structural toxin gene (Hughes *et al*., 1992).

3.2 RtxA can be a novel virulence determinant in weakly pathogenic *Y. enterocolitica* strains

RTX gene cluster encodes the presumptive cytotoxin RtxA, an acyltransferase RtxC, and associated hypothetical proteins Ymp1 and RtxH with unknown functions. RtxA has several motifs, for example hemolysin-type calcium-binding sites, which are bring it closer together with RTX-like cytotoxins. In general, these sites responsible for the hemolysis of erythrocytes. But weakly pathogenic bioserotypes of *Y. enterocolitica* do not possess any hemolytic activity. RtxA has RGD motif that is also present in RtxA from *V. cholerae* but is absent in other RTX toxins. Lally (Lally *et al*., 1997) identified β2 integrin as the cell-surface receptor for the RTX toxins from *Actinobacillus actinomycetemcomitans* and *E. coli*. Given the presence of one RGD motif within RtxA, it may interact with the target cells by binding to host integrins. From the other hand, it has individual distinctive features, which make it unique. For example, glycosaminoglycan attachment site, serralysin-like metalloprotease structure, peptidase C58 *Yersinia*/*Haemophilus* virulence surface antigen domen and prokaryotic membrane lipoprotein lipid attachement site have not been found in other RTX-like toxins.

We speculate based on these data that RtxA from *Y. enterocolitica* may either be a new toxin in the RTX family toxins or a new adhesin. Also we can not exclude a variant when RtxA possesses dual activity as an adhesin and a cytotoxin.

4. pG8786 carries conjugative genes

4.1 pG8786 is an ancient form of the pFra virulence plasmid?

The complete sequence of pG8786, the 137-kb virulence plasmid of the rhamnose-positive *Y. pestis* strain G8786 from a vole (*Microtus arvalis*) in a plague endemic locus in the high mountainous Caucasus in Georgia, was determined. This sequence revealed the recombinant nature of pG8786, namely, the insertion of a 32,617 bp unique *tra* gene cluster and *oriRb* (region 2). Also a larger part of the plasmid pHCM2 of *Salmonella enterica* serovar Typhi that forms the backbone of the *Y. pestis*-specific plasmid pFra is preserved in pG8786 (region 1). However, in contrast to other pFra plasmids (pMT1 from *Y. pestis* biovar Orientalis strain CO92 and pMT1 from *Y. pestis* biovar Mediaevalis strain KIM) the complete region 1 was also found in the pFra plasmid of the avirulent *Y. pestis* biovar Mediaevails strain 91001 isolated from another vole (*Microtus brandti*) in Inner Mongolia, China. The fact that a larger remnant of pHCM2 of the same size is present in both geographically isolated atypical *Y. pestis* strains but absent from the epidemic isolates, implies that the plasmid pFra in *Y. pestis* G8786 and 91001 might represent an ancient form of pFra. This assumption is stressed by the presence of an additional transfer region in pG8786. Alternatively, pG8786 acquired the *tra*-operon more recently by horizontal gene transfer. The variations in the G+C content also point to the chimeric nature of the plasmid. With these facts in mind, one can suppose that pG8786 may have originated by the acquisition of DNA fragments from various microorganisms with a higher G+C content.

4.2 *tra* genes could be acquired due to the IS-mediated recombination events

Comparison of the four sequenced pFra plasmids with pG8786 uncovers extended regions of DNA rearrangements in the backbone of the pFra replicon i.e. large inversions resulting from IS-mediated recombination. Thus, the presence of these flanking IS elements might display DNA rearrangements suffered by the pG8786 plasmid. We suppose that pFra initially co-integrated with a conjugative IncFII group plasmid followed by a subsequent deletion by IS*285*-mediated recombination. Certain *tra*-associated genes (*traM*, *traJ*, *traY*, *traT*, *traS,* and *par*) of pG8786, or the complete *tra* region of another pFra plasmid (pMT1 91001) have suffered such a deletion. Alternatively, the presence of the *par* partition genes, associated with the second RepFII replication origin, might decrease the ability of pG8786 to co-exist with the virulence-associated pYV replicon that belongs probably to the same incompatibility group.

4.3 pG8786 is the potentially transmissive virulence-associated plasmid

Conjugative transfer of the Cm^R-labelled pG8786 could not be demonstrated using the original host *Y. pestis* G8786 and *E. coli* as recipient. Nevertheless, a genetic exchange might take place under certain conditions, possibly in a different ecological environment, e.g. the flea midgut (Hinnebusch *et al*., 2002). Acquisition of a transmissive form of the pFra plasmid (encoding the Fraction 1 antigen and phospholipase D necessary for the colonization of the flea gut) might be the first major step in *Y. pestis* evolution from a common ancestor of *Y. pseudotuberculosis* and *Y. pestis*. Such an acquisition combined with the pre-existing in the genomes of pathogenic *Yersinia* (*Y. pseudotuberculosis*, http://bbrp.llnl.gov/bbrp/html/microbe.html and *Y. enterocolitica*, http://www.sanger.ac.uk/Projects/Y_enterocolitica/) pigmentation locus and insect toxin genes supplied the evolving organism with a better ability to survive in the flea vector resulting in efficient bloodborne transmission. The existence of a potentially transmissive virulence-associated plasmid in *Y. pestis* points to the fact that a new emerging pathogen may appear occasionally with the already acquired ability to survive and multiply efficiently in insect vectors like fleas.

E. SUMMARY

In this study we applied a method of suppression subtractive hybridization (SSH) to map out genomic differences between highly pathogenic *Yersinia enterocolitica* WA-C BG 1B, ST O:8 strain and weakly pathogenic *Y. enterocolitica* Y-108 BG 4, ST O:3 strain (I), as well as Rha-positive *Y. pestis* G8786 bv. Antiqua strain, isolated from a vole in high mountainous Caucasus, Georgia, was compared with another Antiqua isolate, *Y. pestis* Yokohama (II).

I. In total, 428 WA-C-specific and 83 Y-108C-specific sequences were uncovered by SSH applied to the *Y. enterocolitica* isolates. Among them were DNA fragments with similarity to known genes from several groups: (1) genes involved in O-antigen biosynthesis, (2) host-specific restriction-modification systems, (3) systems of iron and heme acquisition and storage, (4) flagellar biogenesis genes, (5) putative virulence factors, (6) drug resistance genes, and (7) mobile elements. In particular, a novel IS-element belonging to the IS*21*-family, designated IS*1331* and the putative toxin RtxA with a high similarity to the RtxA cytotoxin from *V. cholerae* were revealed.

The structure and distribution of IS*1331*, a new *Yersinia enterocolitica* insertion sequence element, were investigated. IS*1331* is related to the IS elements of the IS*21* family and is present in two isoforms in *Y. enterocolitica* ST O:3. The transcription of genes of IS*1331* was shown by RT-PCR and proved its functionality. The probe for IS*1331* efficiently detected all European weakly pathogenic *Y. enterocolitica* bioserotypes, whereas it does not hybridize with other strains. This indicates that IS*1331* can be applied as an additional tool for *Y. enterocolitica* differentiation. Distribution of the different groups of IS elements in two *Y. enterocolitica* pathotypes is in favour of the parallel evolution of American and European *Y. enterocolitica* strains.

We have also characterized the gene cluster in *Y. enterocolitica* O:3 Y-108C that contains four genes of the putative *rtxA* operon: *rtxA*, *rtxC*, *ymp1*, and *rtxH*. The possible toxin, RtxA, resembles members of the RTX (repeats in toxin) toxin family that contains a glycin-rich

repeated motif. Like other RTX toxins, it is associated with an activator, RtxC – acyltransferase. Ymp1, has similarity to a putative membrane protein YPO2693 from *Y. pestis* CO92 and the last one, RtxH, has a similarity to a peptide chain release factor 1 (VC1449) from *V. cholerae*.

Using the Southern blot analysis and DNA microarray we have established that *rtxA* is present only in weakly pathogenic *Y. enterocolitica* subsp. *palearctica* strains and is absent in highly and nonpathogenic bioserotypes. Reverse transcription analysis was carried out to determine the transcription of the genes of the RTX cluster. The positive transcripts indicate in vivo transcription of all four ORFs (*ymp1*, *rtxH*, *rtxC*, and *rtxA*) as a single mRNA. RtxA was immunoprecipitated and subsequently detected using antibody raised against a 79 kDa subfragment of RtxA that was purified from *E. coli* as a 6xHis-tagged protein. Two bands were detected by the RtxA-specific antibody that were not detected by the pre-immune serum. Together these bands consistent with the predicted full-length size of RtxA. These data demonstrate that the RtxA protein is synthesized and exported to culture supernatants but perhaps in a small quantity.

II. Sequences specific for the Rha-positive *Y. pestis* G8786 strain were uncovered by SSH. Several G8786-specific sequences show similarity to genes responsible for the transmissivity of R100 and F plasmids, namely, *traG* (pilus assembly), *traH* (pilus assembly) and *traN* (mating pair stabilization). To map out the location of the insert of the *tra*-genes in the pG8786 plasmid we have applied shotgun sequencing of the complete replicon. The 137,036-bp plasmid pG8786 of *Y. pestis* G8786 isolated in the high mountainous Caucasian plague focus in Georgia is an enlarged form of the pFra virulence-associated plasmid encoding genes for the synthesis of the antigen Fraction 1 and the phospholipase D. In addition to the completely conserved genes of the pFra backbone, pG8786 contains two large regions of 4,642-bp and 32,617-bp, termed regions 1 and 2, respectively. Region 1 retains a larger part of the *Salmonella enterica* serovar Typhi plasmid pHCM2 resembling the backbone of pFra replicons, while region 2 contains 25 ORFs with high similarity to the transfer genes of the F-like plasmids. Despite of the fact that some genes typically involved in conjugative transfer of the F-like replicons are missing in pG8786, we cannot exclude that pG8786 might be transmissive under certain conditions.

This work represents comparative genome analysis of two *Y. enterocolitica* and two *Y. pestis* strains. New unique genetic elements, which are specific for each strain were uncovered. Further intensive experiments have allowed us to characterize a structure and distribution of the new IS element, IS*1331*, a novel putative RtxA-like toxin, and pG8786 plasmid, which carries transfer genes and might be an ancient form of the pFra replicon.

F. REFERENCES

Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., & Carniel, E. (1999) *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A* **96**, 14043-14048.

Al Hasani, K., Rajakumar, K., Bulach, D., Robins-Browne, R., Adler, B., & Sakellaris, H. (2001) Genetic organization of the she pathogenicity island in *Shigella flexneri* 2a. *Microb Pathog* **30**, 1-8.

al Hendy, A., Toivanen, P., & Skurnik, M. (1991) The effect of growth temperature on the biosynthesis of *Yersinia enterocolitica* O:3 lipopolysaccharide: temperature regulates the transcription of the rfb but not of the rfa region. *Microb Pathog* **10**, 81-86.

al Hendy, A., Toivanen, P., & Skurnik, M. (1992) Lipopolysaccharide O side chain of *Yersinia enterocolitica* O:3 is an essential virulence factor in an orally infected murine model. *Infect Immun* **60**, 870-875.

Aleksic, S. (1995) Occurrence of *Y. enterocolitica* antigens O:3, O:9 and O:8 in different *Yersinia* species, their corresponding H antigens and origin. *Contrib Microbiol Immunol* **13**, 89-92.

Aleksic, S. & Bockemuhl, J. (1990) [Microbiology and epidemiology of *Yersinia* infections]. *Immun Infekt* **18**, 178-185.

Alfano, J.R. & Collmer, A. (1997) The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death. *J Bacteriol* **179**, 5655-5662.

Anisimov, A.P., Lindler, L.E., & Pier, G.B. (2004) Intraspecific diversity of *Yersinia pestis*. *Clin Microbiol Rev* **17**, 434-464.

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl. (2000) "Current Protocols in Molecular Biology." John Wiley & Sons, Inc.

Bakanidze, L., Velijanashvili, I., Kekelidze, M., Beridze, L., Zangaladze, E., Zakalashvili, M., Tsereteli, D., & Imnadze, P. (2003) Polymerase chain reaction assays for the presumptive identification of *Yersinia pestis* strains in Georgia. *Adv Exp Med Biol* **529**, 333-336.

Balligand, G., Laroche, Y., & Cornelis, G. (1985) Genetic analysis of virulence plasmid from a serogroup 9 *Yersinia enterocolitica* strain: role of outer membrane protein P1 in resistance to human serum and autoagglutination. *Infect Immun* **48**, 782-786.

Baumann, U., Wu, S., Flaherty, K.M., & McKay, D.B. (1993) Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa*: a two-domain protein with a calcium binding parallel beta roll motif. *EMBO J* **12**, 3357-3364.

Beer, K.B. & Miller, V.L. (1992) Amino acid substitutions in naturally occurring variants of ail result in altered invasion activity. *J Bacteriol* **174**, 1360-1369.

Bengoechea, J.A., Najdenski, H., & Skurnik, M. (2004) Lipopolysaccharide O antigen status of *Yersinia enterocolitica* O:8 is essential for virulence and absence of O antigen affects the expression of other *Yersinia* virulence factors. *Mol Microbiol* **52**, 451-469.

Bengoechea, J.A., Zhang, L., Toivanen, P., & Skurnik, M. (2002) Regulatory network of lipopolysaccharide O-antigen biosynthesis in *Yersinia enterocolitica* includes cell envelope-dependent signals. *Mol Microbiol* **44**, 1045-1062.

Benz, R., Dobereiner, A., Ludwig, A., & Goebel, W. (1992) Haemolysin of *Escherichia coli*: comparison of pore-forming properties between chromosome and plasmid-encoded haemolysins. *FEMS Microbiol Immunol* **5**, 55-62.

Bercovier, H. & Mollaret H.H. (1984) Bergey's manual of systematic bacteriology. Baltimore, Md.: The Williams & Wilkins Co.

Berger, B. & Haas, D. (2001) Transposase and cointegrase: specialized transposition proteins of the bacterial insertion sequence IS*21* and related elements. *Cell Mol Life Sci* **58**, 403-419.

Black, R.E., Jackson, R.J., Tsai, T., Medvesky, M., Shayegani, M., Feeley, J.C., MacLeod, K.I., & Wakelee, A.M. (1978) Epidemic *Yersinia enterocolitica* infection due to contaminated chocolate milk. *N Engl J Med* **298**, 76-79.

Bliska, J.B. & Falkow, S. (1992) Bacterial resistance to complement killing mediated by the Ail protein of *Yersinia enterocolitica*. *Proc Natl Acad Sci U S A* **89**, 3561-3565.

Blomberg, P., Nordstrom, K., & Wagner, E.G. (1992) Replication control of plasmid R1: RepA synthesis is regulated by CopA RNA through inhibition of leader peptide translation. *EMBO J* **11**, 2675-2683.

Bolin, I., Forsberg, A., Norlander, L., Skurnik, M., & Wolf-Watz, H. (1988) Identification and mapping of the temperature-inducible, plasmid-encoded proteins of *Yersinia* spp. *Infect Immun* **56**, 343-348.

Bottone, E.J. (1997) *Yersinia enterocolitica*: the charisma continues. *Clin Microbiol Rev* **10**, 257-276.

Bottone, E.J. (1999) *Yersinia enterocolitica*: overview and epidemiologic correlates. *Microbes Infect* **1**, 323-333.

Boyd A., & Cornelis G.R. (2001) "*Yersinia*". In: Principles of Bacterial Pathogenesis. Ed. by E. Groisman. San Diego: Academic Press, 227-264.

Brown, H.J., Stokes, H.W., & Hall, R.M. (1996) The integrons In*0*, In*2*, and In*5* are defective transposon derivatives. *J Bacteriol* **178**, 4429-4437.

Brubaker, R.R. (1991) Factors promoting acute and chronic diseases caused by yersiniae. *Clin Microbiol Rev* **4**, 309-324.

Brubaker, R.R. (2003) Interleukin-10 and inhibition of innate immunity to *Yersiniae*: roles of Yops and LcrV (V antigen). *Infect Immun* **71**, 3673-3681.

Buchrieser, C., Rusniok, C., Frangeul, L., Couve, E., Billault, A., Kunst, F., Carniel, E., & Glaser, P. (1999) The 102-kilobase pgm locus of *Yersinia pestis*: sequence analysis and comparison of selected regions among different *Yersinia pestis* and *Yersinia pseudotuberculosis* strains. *Infect Immun* **67**, 4851-4861.

Bukholm, G., Kapperud, G., & Skurnik, M. (1990) Genetic evidence that the yopA gene-encoded *Yersinia* outer membrane protein Yop1 mediates inhibition of the anti-invasive effect of interferon. *Infect Immun* **58**, 2245-2251.

Carniel, E. (2001) The *Yersinia* high-pathogenicity island: an iron-uptake island. *Microbes Infect* **3**, 561-569.

Carniel, E. (2002) Plasmids and pathogenicity islands of *Yersinia*. *Curr Top Microbiol Immunol* **264**, 89-108.

Chandler, N.D. & Parisi, M.T. (1994) Radiological case of the month. *Yersinia enterocolitica* masquerading as appendicitis. *Arch Pediatr Adolesc Med* **148**, 527-528.

Chang, Y.F., Young, R., & Struck, D.K. (1989) Cloning and characterization of a hemolysin gene from *Actinobacillus* (*Haemophilus*) *pleuropneumoniae*. *DNA* **8**, 635-647.

Cirillo, S.L., Lum, J., & Cirillo, J.D. (2000) Identification of novel loci involved in entry by *Legionella pneumophila*. *Microbiology* **146 (Pt 6)**, 1345-1359.

Cornelis, G.R. (1994) *Yersinia* pathogenicity factors. *Curr Top Microbiol Immunol* **192**, 243-263.

Cornelis, G.R. (2002) The *Yersinia* Ysc-Yop 'type III' weaponry. *Nat Rev Mol Cell Biol* **3**, 742-752.

Cornelis, G.R., Boland, A., Boyd, A.P., Geuijen, C., Iriarte, M., Neyt, C., Sory, M.P., & Stainier, I. (1998) The virulence plasmid of *Yersinia*, an antihost genome. *Microbiol Mol Biol Rev* **62**, 1315-1352.

Cornelis, G.R., Sluiters, C., Delor, I., Geib, D., Kaniga, K., Lambert, d.R., Sory, M.P., Vanooteghem, J.C., & Michiels, T. (1991) ymoA, a *Yersinia enterocolitica* chromosomal gene modulating the expression of virulence functions. *Mol Microbiol* **5**, 1023-1034.

Cornelis, G.R. & Van Gijsegem, F. (2000) Assembly and function of type III secretory systems. *Annu Rev Microbiol* **54**, 735-774.

Cornelis, G.R. & Wolf-Watz, H. (1997) The *Yersinia* Yop virulon: a bacterial system for subverting eukaryotic cells. *Mol Microbiol* **23**, 861-867.

Cochet, O., Teillaud, J-L., & Sautes, C. (1998) Immunological Techniques Made Easy, John Wiley & Sons. Chichester (eds).

Cover, T.L. & Aber, R.C. (1989) *Yersinia enterocolitica*. *N Engl J Med* **321**, 16-24.

Czuprynski, C.J. & Welch, R.A. (1995) Biological effects of RTX toxins: the possible role of lipopolysaccharide. *Trends Microbiol* **3**, 480-483.

de Almeida, A.M., Guiyoule, A., Guilvout, I., Iteman, I., Baranton, G., & Carniel, E. (1993) Chromosomal irp2 gene in *Yersinia*: distribution, expression, deletion and impact on virulence. *Microb Pathog* **14**, 9-21.

Delor, I. & Cornelis, G.R. (1992) Role of *Yersinia enterocolitica* Yst toxin in experimental infection of young rabbits. *Infect Immun* **60**, 4269-4277.

Delor, I., Kaeckenbeeck, A., Wauters, G., & Cornelis, G.R. (1990) Nucleotide sequence of yst, the *Yersinia enterocolitica* gene encoding the heat-stable enterotoxin, and prevalence of the gene among pathogenic and nonpathogenic yersiniae. *Infect Immun* **58**, 2983-2988.

Deng, W., Burland, V., Plunkett, G., III & other authors (2002) Genome sequence of *Yersinia pestis* KIM. *J Bacteriol* **184**, 4601-4611.

Devignat, R. (1951) Varietes de l'espèce *Pasteurella pestis*: nouvelle hyphothèse. Bull. W. H. O. **4:**247-263.

Diatchenko, L., Lau, Y.F., Campbell, A.P.& other authors (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A* **93**, 6025-6030.

Dower, W.J., Miller, J.F., & Ragsdale, C.W. (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* **16**, 6127-6145.

Ellison, D.W., Lawrenz, M.B., & Miller, V.L. (2004) Invasin and beyond: regulation of *Yersinia* virulence by RovA. *Trends Microbiol* **12**, 296-300.

El Tahir, Y., & Skurnik, M. (2001) YadA, the multifaceted *Yersinia* adhesin. *Int J Med Microbiol* **3**, 209-218.

Fallman, M., Persson, C., & Wolf-Watz, H. (1997) *Yersinia* proteins that target host cell signaling pathways. *J Clin Invest* **99**, 1153-1157.

Felmlee, T., Pellett, S., & Welch, R.A. (1985) Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. *J Bacteriol* **163**, 94-105.

Filippov, A.A., Oleinikov, P.V., Motin, V.L., Protsenko, O.A., & Smirnov, G.B. (1995) Sequencing of two *Yersinia pestis* IS elements, IS*285* and IS*100*. *Contrib Microbiol Immunol* **13**, 306-309.

Filippov, A.A., Solodovnikov, N.S., Kookleva, L.M., & Protsenko, O.A. (1990) Plasmid content in *Yersinia pestis* strains of different origin. *FEMS Microbiol Lett* **55**, 45-48.

Finlay, B.B. & Falkow, S. (1988) Comparison of the invasion strategies used by *Salmonella cholerae-suis*, *Shigella flexneri* and *Yersinia enterocolitica* to enter cultured animal cells: endosome acidification is not required for bacterial invasion or intracellular replication. *Biochimie* **70**, 1089-1099.

Forsberg, A. & Wolf-Watz, H. (1988) The virulence protein Yop5 of *Yersinia pseudotuberculosis* is regulated at transcriptional level by plasmid-plB1-encoded trans-acting elements controlled by temperature and calcium. *Mol Microbiol* **2**, 121-133.

Frey, J. & Kuhnert, P. (2002) RTX toxins in *Pasteurellaceae*. *Int J Med Microbiol* **292**, 149-158.

Frost, L.S., Ippen-Ihler, K., & Skurray, R.A. (1994) Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol Rev* **58**, 162-210.

Galli, D.M., Chen, J., Novak, K.F., & Leblanc, D.J. (2001) Nucleotide sequence and analysis of conjugative plasmid pVT745. *J Bacteriol* **183**, 1585-1594.

Galyov, E.E., Hakansson, S., Forsberg, A., & Wolf-Watz, H. (1993) A secreted protein kinase of *Yersinia pseudotuberculosis* is an indispensable virulence determinant. *Nature* **361**, 730-732.

Gehring, A.M., DeMoll, E., Fetherston, J.D., Mori, I., Mayhew, G.F., Blattner, F.R., Walsh, C.T., & Perry, R.D. (1998) Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem Biol* **5**, 573-586.

Glaser, P., Elmaoglou-Lazaridou, A., Krin, E., Ladant, D., Barzu, O., & Danchin, A. (1989) Identification of residues essential for catalysis and binding of calmodulin in *Bordetella pertussis* adenylate cyclase by site-directed mutagenesis. *EMBO J* **8**, 967-972.

Golubov, A., Heesemann, J., & Rakin, A. (2003) Uncovering genomic differences in human pathogenic *Yersinia enterocolitica*. *FEMS Immunol Med Microbiol* **38**, 107-111.

Golubov, A., Neubauer, H., Nolting, C., Heesemann, J., & Rakin, A. (2004) Structural organization of the pFra virulence-associated plasmid of rhamnose-positive *Yersinia pestis*. *Infect Immun* **72**, 5613-5621.

Grant, T., Bennett-Wood, V., & Robins-Browne, R.M. (1999) Characterization of the interaction between *Yersinia enterocolitica* biotype 1A and phagocytes and epithelial cells in vitro. *Infect Immun* **67**, 4367-4375.

Gripenberg-Lerche, C., Skurnik, M., & Toivanen, P. (1995) Role of YadA-mediated collagen binding in arthritogenicity of *Yersinia enterocolitica* serotype O:8: experimental studies with rats. *Infect Immun* **63**, 3222-3226.

Gutman, L.T., Ottesen, E.A., Quan, T.J., Noce, P.S., & Katz, S.L. (1973) An inter-familial outbreak of *Yersinia enterocolitica* enteritis. *N Engl J Med* **288**, 1372-1377.

Haller, J.C., Carlson, S., Pederson, K.J., & Pierson, D.E. (2000) A chromosomally encoded type III secretion pathway in *Yersinia enterocolitica* is important in virulence. *Mol Microbiol* **36**, 1436-1446.

Han, Y.W. & Miller, V.L. (1997) Reevaluation of the virulence phenotype of the inv yadA double mutants of *Yersinia pseudotuberculosis*. *Infect Immun* **65**, 327-330.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557-580.

Heesemann, J., Algermissen, B., & Laufs, R. (1984) Genetically manipulated virulence of *Yersinia enterocolitica*. *Infect Immun* **46**, 105-110.

Heesemann, J., Hantke, K., Vocke, T., Saken, E., Rakin, A., Stojiljkovic, I., & Berner, R. (1993) Virulence of *Yersinia enterocolitica* is closely associated with siderophore production, expression of an iron-repressible outer membrane polypeptide of 65,000 Da and pesticin sensitivity. *Mol Microbiol* **8**, 397-408.

Hinchliffe, S.J., Isherwood, K.E., Stabler, R.A.& other authors (2003) Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis*. *Genome Res* **13**, 2018-2029.

Hinnebusch, B.J. (2003) Transmission factors: *Yersinia pestis* genes required to infect the flea vector of plague. *Adv Exp Med Biol* **529**, 55-62.

Hinnebusch, B.J., Rudolph, A.E., Cherepanov, P., Dixon, J.E., Schwan, T.G., & Forsberg, A. (2002) Role of *Yersinia* murine toxin in survival of *Yersinia pestis* in the midgut of the flea vector. *Science* **296**, 733-735.

Hoiczyk, E., Roggenkamp, A., Reichenbecher, M., Lupas, A., & Heesemann, J. (2000) Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. *EMBO J* **19**, 5989-5999.

Hu, P., Elliott, J., McCready, P., Skowronski, E., Garnes, J., Kobayashi, A., Brubaker, R.R., & Garcia, E. (1998) Structural organization of virulence-associated plasmids of *Yersinia pestis*. *J Bacteriol* **180**, 5192-5202.

Hueck, C.J. (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* **62**, 379-433.

Hughes, T.R., Piddlesden, S.J., Williams, J.D., Harrison, R.A., & Morgan, B.P. (1992) Isolation and characterization of a membrane protein from rat erythrocytes which inhibits lysis by the membrane attack complex of rat complement. *Biochem J* **284 (Pt 1)**, 169-176.

Iriarte, M. & Cornelis, G.R. (1995) MyfF, an element of the network regulating the synthesis of fibrillae in *Yersinia enterocolitica*. *J Bacteriol* **177**, 738-744.

Iriarte, M., Sory, M.P., Boland, A., Boyd, A.P., Mills, S.D., Lambermont, I., & Cornelis, G.R. (1998) TyeA, a protein involved in control of Yop release and in translocation of *Yersinia* Yop effectors. *EMBO J* **17**, 1907-1918.

Isberg, R.R. (1989) Determinants for thermoinducible cell binding and plasmid-encoded cellular penetration detected in the absence of the *Yersinia pseudotuberculosis* invasin protein. *Infect Immun* **57**, 1998-2005.

Isberg, R.R. & Leong, J.M. (1990) Multiple beta 1 chain integrins are receptors for invasin, a protein that promotes bacterial penetration into mammalian cells. *Cell* **60**, 861-871.

Isberg, R.R., Voorhis, D.L., & Falkow, S. (1987) Identification of invasin: a protein that allows enteric bacteria to penetrate cultured mammalian cells. *Cell* **50**, 769-778.

Iwobi, A., Rakin, A., Garcia, E., & Heesemann, J. (2002) Representational difference analysis uncovers a novel IS*10*-like insertion element unique to pathogenic strains of *Yersinia enterocolitica*. *FEMS Microbiol Lett* **210**, 251-255.

Kapperud, G., Skarpeid, H.J., Solberg, R., & Bergan, T. (1985) Outer membrane proteins and plasmids in different *Yersinia enterocolitica* serogroups isolated from man and animals. *Acta Pathol Microbiol Immunol Scand [B]* **93**, 27-35.

Kay, B.A., Wachsmuth, K., Gemski, P., Feeley, J.C., Quan, T.J., & Brenner, D.J. (1983) Virulence and phenotypic characterization of *Yersinia enterocolitica* isolated from humans in the United States. *J Clin Microbiol* **17**, 128-138.

Kenny, B., Taylor, S., & Holland, I.B. (1992) Identification of individual amino acids required for secretion within the haemolysin (HlyA) C-terminal targeting region. *Mol Microbiol* **6**, 1477-1489.

Kim, S.R., Funayama, N., & Komano, T. (1993) Nucleotide sequence and characterization of the *traABCD* region of IncI1 plasmid R64. *J Bacteriol* **175**, 5035-5042.

Kinder, S.A., Badger, J.L., Bryant, G.O., Pepe, J.C., & Miller, V.L. (1993) Cloning of the *Yen*I restriction endonuclease and methyltransferase from *Yersinia enterocolitica* serotype O8 and construction of a transformable R-M+ mutant. *Gene* **136**, 271-275.

Koebnik, R., Hantke, K., & Braun, V. (1993) The TonB-dependent ferrichrome receptor FcuA of *Yersinia enterocolitica*: evidence against a strict co-evolution of receptor structure and substrate specificity. *Mol Microbiol* **7**, 383-393.

Kutyrev, V.V., Popov, I., & Protsenko, O.A. (1986) [Plasmids of the pathogenicity of *Yersinia pestis*]. *Mol Gen Mikrobiol Virusol* 3-11.

Lahteenmaki, K., Virkola, R., Saren, A., Emody, L., & Korhonen, T.K. (1998) Expression of plasminogen activator pla of *Yersinia pestis* enhances bacterial attachment to the mammalian extracellular matrix. *Infect Immun* **66**, 5755-5762.

Lally, E.T., Kieba, I.R., Sato, A.& other authors (1997) RTX toxins recognize a beta2 integrin on the surface of human target cells. *J Biol Chem* **272**, 30463-30469.

Lawley, T.D., Klimke, W.A., Gubbins, M.J., & Frost, L.S. (2003) F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett* **224**, 1-15.

Lee, L.A., Gerber, A.R., Lonsway, D.R.& other authors (1990) *Yersinia enterocolitica* O:3 infections in infants and children, associated with the household preparation of chitterlings. *N Engl J Med* **322**, 984-987.

Lian, C.J. & Pai, C.H. (1985) Inhibition of human neutrophil chemiluminescence by plasmid-mediated outer membrane proteins of *Yersinia enterocolitica*. *Infect Immun* **49**, 145-151.

Lin, W., Fullner, K.J., Clayton, R., Sexton, J.A., Rogers, M.B., Calia, K.E., Calderwood, S.B., Fraser, C., & Mekalanos, J.J. (1999) Identification of a vibrio cholerae RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. *Proc Natl Acad Sci U S A* **96**, 1071-1076.

Linde, H.J., Neubauer, H., Meyer, H., Aleksic, S., & Lehn, N. (1999) Identification of *Yersinia* species by the Vitek GNI card. *J Clin Microbiol* **37**, 211-214.

Lindler, L.E., Plano, G.V., Burland, V., Mayhew, G.F., & Blattner, F.R. (1998) Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIM5 plasmid encoding murine toxin and capsular antigen. *Infect Immun* **66**, 5731-5742.

Lindler, L. E., & Tall, B. D. (1993) *Yersinia pestis* pH 6 antigen forms fimbriae and is induced by intracellular association with macrophages. *Mol Microbiol* **8**, 311-324.

Lisitsyn, N. & Wigler, M. (1995) Representational difference analysis in detection of genetic lesions in cancer. *Methods Enzymol* **254**, 291-304.

Lisitsyn, N.A., Rosenberg, M.V., Launer, G.A., Wagner, L.L., Potapov, V.K., Kolesnik, T.B., & Sverdlov, E.D. (1993) A method for isolation of sequences missing in one of two related genomes. *Mol Gen Mikrobiol Virusol* 26-29.

Lo, R.Y., Strathdee, C.A., & Shewen, P.E. (1987) Nucleotide sequence of the leukotoxin genes of *Pasteurella haemolytica* A1. *Infect Immun* **55**, 1987-1996.

Lockman, H.A., Gillespie, R.A., Baker, B.D., & Shakhnovich, E. (2002) *Yersinia pseudotuberculosis* produces a cytotoxic necrotizing factor. *Infect Immun* **70**, 2708-2714.

Lohe, A.R., Sullivan, D.T., & Hartl, D.L. (1996) Subunit interactions in the mariner transposase. *Genetics* **144**, 1087-1095.

Lu, J., Manchak, J., Klimke, W., Davidson, C., Firth, N., Skurray, R.A., & Frost, L.S. (2002) Analysis and characterization of the IncFV plasmid pED208 transfer region. *Plasmid* **48**, 24-37.

Mack, D., Heesemann, J., & Laufs, R. (1994) Characterization of different oligomeric species of the *Yersinia enterocolitica* outer membrane protein YadA. *Med Microbiol Immunol (Berl)* **183**, 217-227.

Mahillon, J. & Chandler, M. (1998) Insertion sequences. *Microbiol Mol Biol Rev* **62**, 725-774.

Makino, K., Yokoyama, K., Kubota, Y.& other authors (1999) Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak. *Genes Genet Syst* **74**, 227-239.

Martinevskii, I. L. (1969) Biology and genetic features of plague and plague-related microbes. Meditsina Press, Moscow.

Michiels, T., Vanooteghem, J.C., Lambert, d.R., China, B., Gustin, A., Boudry, P., & Cornelis, G.R. (1991) Analysis of *virC*, an operon involved in the secretion of Yop proteins by *Yersinia enterocolitica*. *J Bacteriol* **173**, 4994-5009.

Miller, V.L., Bliska, J.B., & Falkow, S. (1990) Nucleotide sequence of the *Yersinia enterocolitica ail* gene and characterization of the Ail protein product. *J Bacteriol* **172**, 1062-1069.

Miller, V.L. & Falkow, S. (1988) Evidence for two genetic loci in *Yersinia enterocolitica* that can promote invasion of epithelial cells. *Infect Immun* **56**, 1242-1248.

Miller, V.L., Farmer, J.J., III, Hill, W.E., & Falkow, S. (1989) The *ail* locus is found uniquely in *Yersinia enterocolitica* serotypes commonly associated with disease. *Infect Immun* **57**, 121-131.

Mills, M. & Payne, S.M. (1995) Genetics and regulation of heme iron transport in *Shigella dysenteriae* and detection of an analogous system in *Escherichia coli* O157:H7. *J Bacteriol* **177**, 3004-3009.

Mills, S.D., Boland, A., Sory, M.P., van der, S.P., Kerbourch, C., Finlay, B.B., & Cornelis, G.R. (1997) *Yersinia enterocolitica* induces apoptosis in macrophages by a process requiring functional type III secretion and translocation mechanisms and involving YopP, presumably acting as an effector protein. *Proc Natl Acad Sci U S A* **94**, 12638-12643.

Motin, V.L., Georgescu, A.M., Elliott, J.M. & other authors (2002) Genetic variability of *Yersinia pestis* isolates as predicted by PCR-based IS*100* genotyping and analysis of structural genes encoding glycerol-3-phosphate dehydrogenase (*glpD*). *J Bacteriol* **184**, 1019-1027.

Murphy, K.C. (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. *J Bacteriol* **180**, 2063-2071.

Neubauer, H., Aleksic, S., Hensel, A., Finke, E.J., & Meyer, H. (2000) *Yersinia enterocolitica* 16S rRNA gene types belong to the same genospecies but form three homology groups. *Int J Med Microbiol* **290**, 61-64.

Neubauer, H., Reischl, U., Kostler, J., Aleksic, S., Finke, E.J., & Meyer, H. (1999) Variations in the 16S rRNA gene sequence of *Yersinia enterocolitica* isolates influence the specificity of molecular identification systems. *Zentralbl Bakteriol* **289**, 329-337.

Neubauer, H., Sauer, T., Becker, H., Aleksic, S., & Meyer, H. (1998) Comparison of systems for identification and differentiation of species within the genus *Yersinia*. *J Clin Microbiol* **36**, 3366-3368.

Neubauer, H., Sprague, L.D., Hensel, A., Aleksic, S., & Meyer, H. (2000) Specific detection of plasmid bearing *Yersinia* isolates by PCR. *Clin Lab* **46**, 583-587.

Nikolaus, T., Deiwick, J., Rappl, C., Freeman, J.A., Schroder, W., Miller, S.I., & Hensel, M. (2001) SseBCD proteins are secreted by the type III secretion system of *Salmonella* pathogenicity island 2 and function as a translocon. *J Bacteriol* **183**, 6036-6045.

Osborn, A.M., Silva Tatley, F.M., Steyn, L.M., Pickup, R.W., & Saunders, J.R. (2000) Mosaic plasmids and mosaic replicons: evolutionary lessons from the analysis of genetic diversity in IncFII-related replicons. *Microbiology* **146 (Pt 9)**, 2267-2275.

Pansegrau, W., Lanka, E., Barth, P.T. & other authors (1994) Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. *J Mol Biol* **239**, 623-663.

Park, H.I. & Ming, L.J. (2002) Mechanistic studies of the astacin-like *Serratia* metalloendopeptidase serralysin: highly active (>2000%) Co(II) and Cu(II) derivatives for further corroboration of a "metallotriad" mechanism. *J Biol Inorg Chem* **7**, 600-610.

Parkhill, J., Dougan, G., James, K.D. & other authors (2001) Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**, 848-852.

Parkhill, J., Wren, B.W., Thomson, N.R. & other authors (2001) Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523-527.

Pellett, S. & Welch, R.A. (1996) *Escherichia coli* hemolysin mutants with altered target cell specificity. *Infect Immun* **64**, 3081-3087.

Perry, R.D. & Fetherston, J.D. (1997) *Yersinia pestis* - etiologic agent of plague. *Clin Microbiol Rev* **10**, 35-66.

Petersen, S. & Young, G.M. (2002) Essential role for cyclic AMP and its receptor protein in *Yersinia enterocolitica* virulence. *Infect Immun* **70**, 3665-3672.

Pierson, D.E. & Falkow, S. (1993) The ail gene of *Yersinia enterocolitica* has a role in the ability of the organism to survive serum killing. *Infect Immun* **61**, 1846-1852.

Podladchikova, O.N., Dikhanov, G.G., Rakin, A.V., & Heesemann, J. (1994) Nucleotide sequence and structural organization of *Yersinia pestis* insertion sequence IS*100*. *FEMS Microbiol Lett* **121**, 269-274.

Posey, J.E. & Gherardini, F.C. (2000) Lack of a role for iron in the Lyme disease pathogen. *Science* **288**, 1651-1653.

Praszkier, J., Wei, T., Siemering, K., & Pittard, J. (1991) Comparative analysis of the replication regions of IncB, IncK, and IncZ plasmids. *J Bacteriol* **173**, 2393-2397.

Protsenko, O.A., Filippov, A.A., & Kutyrev, V.V. (1991) Integration of the plasmid encoding the synthesis of capsular antigen and murine toxin into *Yersinia pestis* chromosome. *Microb Pathog* **11**, 123-128.

Radnedge, L., Agron, P.G., Worsham, P.L., & Andersen, G.L. (2002) Genome plasticity in *Yersinia pestis*. *Microbiology* **148**, 1687-1698.

Radnedge, L., Gamez-Chin, S., McCready, P.M., Worsham, P.L., & Andersen, G.L. (2001) Identification of nucleotide sequences for the specific and rapid detection of *Yersinia pestis*. *Appl Environ Microbiol* **67**, 3759-3762.

Rakin, A. & Heesemann, J. (1995) Virulence-associated fyuA/irp2 gene cluster of *Yersinia enterocolitica* biotype 1B carries a novel insertion sequence IS*1328*. *FEMS Microbiol Lett* **129**, 287-292.

Rakin, A., Schubert, S., Guilvout, I., Carniel, E., & Heesemann, J. (2000) Local hopping of IS*3* elements into the A+T-rich part of the high-pathogenicity island in *Yersinia enterocolitica* 1B, O:8. *FEMS Microbiol Lett* **182**, 225-229.

Revell, P.A. & Miller, V.L. (2000) A chromosomally encoded regulator is required for expression of the *Yersinia enterocolitica* inv gene and for virulence. *Mol Microbiol* **35**, 677-685.

Rohde, J.R., Fox, J.M., & Minnich, S.A. (1994) Thermoregulation in *Yersinia enterocolitica* is coincident with changes in DNA supercoiling. *Mol Microbiol* **12**, 187-199.

Rosenshine, I., Duronio, V., & Finlay, B.B. (1992) Tyrosine protein kinase inhibitors block invasin-promoted bacterial uptake by epithelial cells. *Infect Immun* **60**, 2211-2217.

Rosqvist, R., Skurnik, M., & Wolf-Watz, H. (1988) Increased virulence of *Yersinia pseudotuberculosis* by two independent mutations. *Nature* **334**, 522-524.

Rossi, M.S., Fetherston, J.D., Letoffe, S., Carniel, E., Perry, R.D., & Ghigo, J.M. (2001) Identification and characterization of the hemophore-dependent heme acquisition system of *Yersinia pestis*. *Infect Immun* **69**, 6707-6717.

Ruoslahti, E. (1996) RGD and other recognition sequences for integrins. *Annu Rev Cell Dev Biol* **12**, 697-715.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., & Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.

Saken, E., Roggenkamp, A., Aleksic, S., & Heesemann, J. (1994) Characterisation of pathogenic *Yersinia enterocolitica* serogroups by pulsed-field gel electrophoresis of genomic *Not*I restriction fragments. *J Med Microbiol* **41**, 329-338.

Schiemann, D.A. (1982) Development of a two-step enrichment procedure for recovery of *Yersinia enterocolitica* from food. *Appl Environ Microbiol* **43**, 14-27.

Schubert, S., Fischer, D., & Heesemann, J. (1999) Ferric enterochelin transport in *Yersinia enterocolitica*: molecular and evolutionary aspects. *J Bacteriol* **181**, 6387-6395.

Schubert, S., Bockemuhl, J., Brendler, U., & Heesemann, J. (2003) First isolation of virulent *Yersinia enterocolitica* O8, biotype 1B in Germany. Eur J Clin Microbiol Infect Dis **22**, 66-68.

Schulte, R., Wattiau, P., Hartland, E.L., Robins-Browne, R.M., & Cornelis, G.R. (1996) Differential secretion of interleukin-8 by human epithelial cell lines upon entry of virulent or nonvirulent *Yersinia enterocolitica*. *Infect Immun* **64**, 2106-2113.

Seoane, A. & Garcia Lobo, J.M. (2000) Identification of a streptogramin A acetyltransferase gene in the chromosome of *Yersinia enterocolitica*. *Antimicrob Agents Chemother* **44**, 905-909.

Skurnik, M. (1984) Lack of correlation between the presence of plasmids and fimbriae in *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*. *J Appl Bacteriol* **56**, 355-363.

Skurnik, M., Bolin, I., Heikkinen, H., Piha, S., & Wolf-Watz, H. (1984) Virulence plasmid-associated autoagglutination in *Yersinia* spp. *J Bacteriol* **158**, 1033-1036.

Skurnik, M., Peippo, A., & Ervela, E. (2000) Characterization of the O-antigen gene clusters of *Yersinia pseudotuberculosis* and the cryptic O-antigen gene cluster of *Yersinia pestis* shows that the plague bacillus is most closely related to and has evolved from *Y. pseudotuberculosis* serotype O:1b. *Mol Microbiol* **37**, 316-330.

Skurnik, M. & Wolf-Watz, H. (1989) Analysis of the *yopA* gene encoding the Yop1 virulence determinants of *Yersinia* spp. *Mol Microbiol* **3**, 517-529.

Som, S., Bhagwat, A.S., & Friedman, S. (1987) Nucleotide sequence and expression of the gene encoding the *Eco*RII modification enzyme. *Nucleic Acids Res* **15**, 313-332.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**, 503-517.

Stover, C.K., Pham, X.Q., Erwin, A.L. & other authors (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406**, 959-964.

Straley, S.C. & Perry, R.D. (1995) Environmental modulation of gene expression and pathogenesis in *Yersinia*. *Trends Microbiol* **3**, 310-317.

Strauch, E., Goelz, G., Knabner, D., Konietzny, A., Lanka, E., & Appel, B. (2003) A cryptic plasmid of *Yersinia enterocolitica* encodes a conjugative transfer system related to the regions of CloDF13 Mob and IncX Pil. *Microbiology* **149**, 2829-2845.

Tamm, A., Tarkkanen, A.M., Korhonen, T.K., Kuusela, P., Toivanen, P., & Skurnik, M. (1993) Hydrophobic domains affect the collagen-binding specificity and surface polymerization as well as the virulence potential of the YadA protein of *Yersinia enterocolitica*. *Mol Microbiol* **10**, 995-1011.

Tauxe, R.V., Vandepitte, J., Wauters, G., Martin, S.M., Goossens, V., De Mol, P., Van Noyen, R., & Thiers, G. (1987) *Yersinia enterocolitica* infections and pork: the missing link. *Lancet* **1**, 1129-1132.

Tertti, R., Skurnik, M., Vartio, T., & Kuusela, P. (1992) Adhesion protein YadA of *Yersinia* species mediates binding of bacteria to fibronectin. *Infect Immun* **60**, 3021-3024.

Thompson, S.A., Wang, L.L., West, A., & Sparling, P.F. (1993) *Neisseria meningitidis* produces iron-regulated proteins related to the RTX family of exoproteins. *J Bacteriol* **175**, 811-818.

Towbin, H., Staehelin, T., & Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A* **76**, 4350-4354.

Trebesius, K., Harmsen, D., Rakin, A., Schmelz, J., & Heesemann,J. (1998) Development of rRNA-targeted PCR and in situ hybridization with fluorescently labelled oligonucleotides for detection of *Yersinia* species. *J Clin Microbiol* **36**, 2557-2564.

Worsham, P.L. & Roy, C. (2003) Pestoides F, a *Yersinia pestis* strain lacking plasminogen activator, is virulent by the aerosol route. *Adv Exp Med Biol* **529**, 129-131.

Yang, Y., Merriam, J.J., Mueller, J.P., & Isberg, R.R. (1996) The *psa* locus is responsible for thermoinducible binding of *Yersinia pseudotuberculosis* to cultured cells. *Infect Immun* **64**, 2483-2489.

Yanisch-Perron, C., Vieira, J., & Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

Young, V.B., Falkow, S., & Schoolnik, G.K. (1992) The invasin protein of *Yersinia enterocolitica*: internalization of invasin-bearing bacteria by eukaryotic cells is associated with reorganization of the cytoskeleton. *J Cell Biol* **116**, 197-207.

Zhang, L., al Hendy, A., Toivanen, P., & Skurnik, M. (1993) Genetic organization and sequence of the rfb gene cluster of *Yersinia enterocolitica* serotype O:3: similarities to the dTDP-L-rhamnose biosynthesis pathway of *Salmonella* and to the bacterial polysaccharide transport systems. *Mol Microbiol* **9**, 309-321.

Zhang, L., Radziejewska-Lebrecht, J., Krajewska-Pietrasik, D., Toivanen, P., & Skurnik, M. (1997) Molecular and chemical characterization of the lipopolysaccharide O-

antigen and its role in the virulence of *Yersinia enterocolitica* serotype O:8. *Mol Microbiol* **23**, 63-76.

Zhang, Y.L., Ong, C.T., & Leung, K.Y. (2000) Molecular analysis of genetic differences between virulent and avirulent strains of *Aeromonas hydrophila* isolated from diseased fish. *Microbiology* **146 (Pt 4)**, 999-1009.

G. ABBREVIATIONS

dest	distilled water
Amp	Ampicillin
AP	Alkaline Phosphatase
APS	Ammonium persulphate
ATP	Adenosine Triphosphate
BCIP	5-Bromo-6-chloro-3-indolylphosphate
BG	Biogroup
bp	base pair
BSA	Bovine serum albumen
CIAP	Calf intestinal alkaline phosphatase
Cm	Chloramphenicol
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate (dATP, dCTP, dGTP, dTTP)
dsDNA	double stranded DNA
EDTA	Ethylenediamine triacetic acid
e.g.	for example
EtOH	Ethanol
Fig.	Figure
hr	hour
HPI	High pathogenicity island
IPTG	Isopropyl ß-D-thiogalactopyranoside
kb	kilobase
Kan	Kanamycin
kDa	kilo Dalton
LB	Luria Bertani
µ	micro
m	milli
M	molar
MCS	multiple cloning site
min	minute
n	nano

Nal	Nalidixic acid
NBT	Nitroblue tetrazolium
OD	optical density
o/n	overnight
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	polymerase chain reaction
Rha	rhamnose
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	revolutions per minute
RT	Room temperature
RT	Reverse Transcriptase / Reverse transcription
SDS	Sodium dodecyl sulphate
sec	seconds
ssDNA	single stranded DNA
Tab.	Table
TAE	Tris-Acetate-EDTA
TEMED	N,N,N’,N’-tetramethyl-ethylenediamine
Tet	Tetracycline
TRIS	Tris-(hydroxymethyl)-ammonium methane
O/N	overnight
UV	Ultraviolet
V	Volt
Vol.	Volume
v/v	volume by volume
wt	wild type
w/v	weight by volume
X-gal	5-bromo-4-chloro-3-indolyl-ß-galactoside

Nucleic acids

A: Adenine

C: Cytosine

G: Guanine

T: Thymine

PUBLISHED ASPECTS OF THIS WORK

Golubov A, Neubauer H, Nolting C, Heesemann J, and Rakin A. 2004. Structural organization of the pFra virulence-associated plasmid of rhamnose-positive *Yersinia pestis*. Infect Immun., **72**: 5613-21.

Golubov, A., Heesemann, J., and Rakin, A. 2003. Uncovering genomic differences in human pathogenic *Yersinia enterocolitica*. FEMS Immunology and Medical Microbiology, **38**: 107-111.

Rakin A., Golubov A., Iwobi A., and Heesemann J. 2003. Tracing Acquisitions and Losses in *Yersinia* Genomes. In “The Genus *Yersinia*: Entering the Functional Genomics Era”. Proceedings of the 8th International conference on *Yersinia.* (Skurnik, M., Bengoechea, J. A., and Granfors, K., Eds.). Turku, Finland.

CURRICULUM VITAE

Andrey Golubov

ACADEMIC BACKGROUND AND QUALIFICATIONS:

08/2001 – 12/2004	**Doctorate research work** at Max von Pettenkofer Institute for Hygiene and Medical Microbiology of the Ludwig Maximilians University (LMU), Munich, Germany. Research theme: Comparative genome analysis of *Yersinia*
1995-2001	**M.Sc. General Biology/Biochemistry**; Rostov State University, Rostov-on-Don, Russia. M.Sc. Thesis: Identification, cloning and characterization of a novel genes involved in production of exopolysaccharides by *Rhizobium leguminosarum* bv. *Viciae* VF39
1992-1995	**B.Sc. (High distinction) Medicine;** Diploma of a Medical Assistent, Shakhtinskoye Medical School, Shakhty, Russia.

AFFILIATIONS:

Member of the European Federation of Biotechnology
Ms. Sylvia Oreifig (Acting General Secretary)
Poortlandplein 6 NL-2628 BM Delft

Personal Data

Date of Birth:	October 03^{d}, 1975
Place of Birth:	Shakhty, Russia